E C O N O M I C S B U L L E T I N

THE DETERMINANTS OF COSTS AND EFFICIENCIES WHERE PRODUCERS ARE HETEROGENEOUS: THE CASE OF SPANISH UNIVERSITIES

Geraint Johnes
Lancaster University Management School

Manuel Salas-Velasco
Universidad de Granada

Abstract

A multi-product cost function is evaluated for the universities of Spain, using a random parameters stochastic frontier model. This allows estimates of systematic cost differences to be obtained alongside estimates of universities' efficiency. In addition, we evaluate average incremental costs of key university output, and provide measures of economies of scale and scope.

Citation: Johnes, Geraint and Manuel Salas-Velasco, (2007) "THE DETERMINANTS OF COSTS AND EFFICIENCIES WHERE PRODUCERS ARE HETEROGENEOUS: THE CASE OF SPANISH UNIVERSITIES." *Economics Bulletin*, Vol. 4, No. 15 pp. 1-9
Submitted: April 24, 2007. **Accepted:** April 24, 2007.
URL: http://economicsbulletin.vanderbilt.edu/2007/volume4/EB-07D20004A.pdf

1. Introduction

The literature on empirical cost functions has frequently appealed to universities as a source of data. Universities are multi-product organisations, and in many countries data on each of their outputs (including teaching of various types and research) are fairly easy to come by. Likewise, data on their costs are published in many countries. The literature on higher education cost functions has therefore grown to be quite large; examples include Cohn *et al.* (1989), Johnes (1996, 1997, 1998), Izadi *et al.* (2002), Johnes *et al.* (2005), and Stevens (2005). It is not surprising, then, to note that many of the advances that have been made to the literature on empirical cost functions have been made using data on higher education institutions. For example, over the last decade or so, it has become the norm to estimate such functions using frontier methods (since Johnes, 1996). In this paper, we extend the literature further by estimating a random parameter stochastic frontier cost function for the universities of Spain. This allows us simultaneously to evaluate efficiency, using the method of Jondrow *et al.* (1982) and the extent to which the underlying cost structures differ across institutions.

The paper has the following structure. In the next section we outline the methodology. The third section discusses data issues, while the fourth focuses on our results. A fifth section draws together the conclusions of the paper.

2. Methodology

The conventional approach to stochastic frontier estimation, based upon cross-section data, is due to Aigner *et al.* (1977). In this model, the equation

$$y_i = \alpha + \boldsymbol{\beta}'\mathbf{x}_i + v_i \pm u_i \tag{1}$$

is estimated using maximum likelihood, where y_i denotes the dependent variable (typically costs or output) for the ith unit of observation, $\mathbf{x}_i$ is a vector of explanatory variables, v_i denotes normally distributed white noise, typically attributed to measurement error, and u_i is a second residual term that is intended to capture efficiency differences across observations. This last term is added to (subtracted from) the other terms on the right hand side of (1) if a cost (production) function is being estimated. The residuals v and u may be aggregated to give the total regression residual, ε. The u component of the residual could in principle follow any non-normal distribution (so that it can be distinguished from v); for reasons of analytical convenience the half-normal is a common assumption, and that is what we assume in the sequel.

Following the insight of Jondrow *et al.* (1982) it is possible to recover observation-specific estimates of the efficiency residual. This is given by

$$\mathrm{E}[u_i|\varepsilon_i] = \sigma\lambda\{\phi(a_i)/[1-\Phi(a_i)] - a_i\}/(1+\lambda^2) \tag{2}$$

where $\sigma = (\sigma_v^2 + \sigma_u^2)^{1/2}$, $\lambda = \sigma_u / \sigma_v$, $a_i = \pm\varepsilon_i\lambda/\sigma$, and $\phi(.)$and $\Phi(.)$ are, respectively, the density and distribution of the standard normal.

The major innovation of the present paper is to apply frontier methods in the context of panel data. In this case it is appropriate to modify (1) to

$$y_{it} = \alpha_i + \boldsymbol{\beta}'_i\,\mathbf{x}_{it} + v_{it} \pm u_{it} \tag{3}$$

where $v_{it} \sim N[0, \sigma_v^2]$, $u_{it} = |U_{it}|$, $U_{it} \sim N[0, \sigma_{ui}^2]$, and v_{it} is independent of u_{it}. Equation (2) is likewise modified, for the panel data case, to

$$\mathrm{E}[u_{it}|\varepsilon_{\mathrm{it}}] = \sigma\lambda\{\phi(a_{it})/[1-\Phi(a_{it})] - a_{it}\}/(1+\lambda^2) \qquad (4)$$

We model the β_i of equation (3) as random parameters. Greene (2005) summarises the problem by defining the stochastic frontier as in (3) above, the inefficiency distribution as a half-normal with mean $\mu_i = \boldsymbol{\mu}'_i\mathbf{z}_i$ and standard deviation $\sigma_{ui} = \sigma_u\exp(\theta'_i\mathbf{h}_i)$, and the parameter heterogeneity as follows:

$$\left.\begin{aligned}
(\alpha_i, \boldsymbol{\beta}_i) &= (\bar{\alpha}, \bar{\boldsymbol{\beta}}) + \Delta_{\alpha,\beta}\mathbf{q}_i + \boldsymbol{\Gamma}_{\alpha,\boldsymbol{\beta}}\mathbf{w}_{\alpha,\boldsymbol{\beta}i} \\
\boldsymbol{\mu}_i &= \bar{\boldsymbol{\mu}} + \boldsymbol{\Delta}_{\boldsymbol{\mu}}\mathbf{q}_i + \boldsymbol{\Gamma}_{\boldsymbol{\mu}}\mathbf{w}_{\mu i} \\
\boldsymbol{\theta}_i &= \bar{\boldsymbol{\theta}} + \boldsymbol{\Delta}_{\boldsymbol{\theta}}\mathbf{q}_i + \boldsymbol{\Gamma}_{\boldsymbol{\theta}}\mathbf{w}_{\boldsymbol{\theta}i}
\end{aligned}\right\} \qquad (5)$$

Here the random variation appears in the random parameters vector $\mathbf{w}_{ji}$ (where i is the index of producers and j refers to either the constant, the slope parameter, or – in more general specifications of the model - the moments of the inefficiency distribution represented by μ and θ); this vector is here assumed to have mean vector zero and, in the case where (as here) parameters are assumed to be normally distributed, the covariance matrix equals the identity matrix.

The parameters of this model must, owing to the presence of an unclosed integral in the unconditional log likelihood, be estimated by maximisation of the simulated log likelihood function:

$$\log L_S = \sum_{i=1}^{N}\frac{1}{R}\sum_{r=1}^{R}\{\sum_{t=1}^{T}\ln\Phi\{[\mu_{ir}/(\sigma_{uir}/\sigma_v) \pm (y_{it} - \alpha_{ir} - \boldsymbol{\beta}'_{ir}\mathbf{x}_{it})(\sigma_{uir}/\sigma_v)]/\sqrt{\sigma_{uir}^2 + \sigma_v^2}\} -$$
$$\frac{1}{2}\{[\mu_i \pm (y_{it} - \alpha_{ir} - \boldsymbol{\beta}'_{ir}\mathbf{x}_{it})]/\sqrt{\sigma_{uir}^2 + \sigma_v^2}\}^2 + \ln\frac{1}{\sqrt{2\pi}} - \ln\Phi(\mu_i/\sigma_{uir}) - \ln\sqrt{\sigma_{uir}^2 + \sigma_v^2}\} \quad (6)$$

The model is estimated using Limdep.

3. Data

Data on the public universities of Spain are published biennially by the Conferencia de Rectores de las Universidades Españolas (CRUE) and are available at the website http://www.crue.org. In this paper we use data for individual institutions in 1998, 2000, 2002 and 2004 to form a panel. Several institutions have incomplete data, and these are excluded from the sample used in the present exercise; the excluded universities do not appear in any way to be systematically distinct from those that are retained in the dataset, and we do not therefore believe that the question of data availability introduces any estimation bias due to selection of the sample.

Student numbers data are disaggregated to broad subject area. Two subjects are defined in the present study – non-science and science. Early experimentation showed that further disaggregation causes problems of multicollinearity, particularly between variables denoting numbers of undergraduate students in different subjects. Data on mode of study (part-time versus full-time) are not available, since the vast majority of Spanish students are studying on a full-time basis.

The data on costs are published as expenditure per undergraduate student. These have therefore been multiplied by the total number of such students, and then deflated to year 2000 prices, to arrive at the figures used here.

The CRUE data provide information about a variety of measures that could be used for the research variable. In common with earlier studies in this area, we use research funding as an indicator of research activity. This is preferred to data on publications (available from the Institute for Scientific Information), since it provides a measure of the value that the market places on research done in the various institutions. It may therefore be regarded as a quality-adjusted measure of the quantity of research that is being undertaken.

Descriptive statistics appear in Table 1. The typical Spanish university is quite large in comparison with those in several other European countries, with about 20000 students studying for degrees below doctoral level. One reason for this is that the duration of studies has been long in comparison with the norm in Anglo-Saxon higher education systems. This will change over the coming years as Spain's universities amend their provision in line with the requirements of the Bologna Accord, which aims to harmonise higher education systems in some 40 European countries by the year 2010.

4. Results

The results of our regression analysis are reported in Table 2. The functional form of the cost function is quadratic, in line with the work of Baumol *et al*. (1982) on multi-product cost functions, and in line also with the body of empirical work done over recent years on university cost functions – see for example Cohn *et al*. (1989), Johnes *et al*. (2005). The model is estimated as a random parameters stochastic frontier, in which the residual is decomposed into a normal and a half-normal component (Tsionas, 2002; Greene, 2005). The speed of the solution has been increased by using Halton (1960) sequences. A random parameter attaches to the constant in the equation. If this were the only random parameter, the equation would be a random effects model. But we also attach a random parameter to the research variable. Extensive experimentation indicated that this is the only variable (other than the constant) where the randomisation of the parameter improves the fit provided by the equation. In the case of both the constant and the research variable, the distribution followed by the random parameter is normal.

In view of the highly nonlinear specification, the coefficients of this model are difficult to interpret. Note, however, that it is possible to obtain from these coefficients, and from the descriptive statistics, a large amount of information about unit costs and returns to scale and scope. These will be discussed later.

In the meantime, it is instructive to examine the information about institution-specific parameters and efficiency that is provided by the model. Table 3 provides data, separately for each institution in the sample, on the random parameters and the efficiency measure obtained from the half-normal residual. The latter measure is obtained by dividing the predicted value of costs on the frontier by the predicted value of costs plus the one-sided residual; hence the efficiency is defined to lie within the unit interval, with a value of one representing an institution that lies on the efficiency frontier. It is readily observed that all universities in the sample achieve extremely high levels of efficiency. Indeed, the insignificant values of λ and σ in Table 2 suggest that efficiencies are generally so high that the frontier model is insignificantly different from a least squares random parameter specification. There is, however, considerable variation in both fixed costs and in the cost of research across institutions. Fixed costs are unusually high at Salamanca (the most ancient of the Spanish universities), Pompeu Fabra, Las Palmas de Gran Canaria and Pais Vasco. The first of these is an ancient university with high costs attached to the maintenance of buildings; the second is a multicampus institution located in a large conurbation; costs in the third are likely affected by its island location; finally, Pais Vasco has three campuses that are located at a distance from one another. Fixed costs are unusually low at Jaen, a recently opened single campus institution located on the outskirts of the town. Costs of producing research are unusually high at Salamanca, Castilla La Mancha, and Pais Vasco. The high costs of research in the last of these are likely due to the organisation of discipline-specific activity across the institution's campuses.

Average incremental costs, evaluated at the means of the random parameters, are reported in Table 4 for the case of a typical university – that is, one that produces mean quantities of all outputs. These are obtained using the method devised by Baumol *et al*. (1982) and widely used elsewhere in the literature on university cost functions. In common with findings from other countries, science tuition is more costly to the typical university than non-science tuition. The cost of producing postgraduate students is higher than that of producing undergraduates in any subject area, again a finding common to studies done in other countries. The average incremental cost of research is high, suggesting that each euro of additional research funding adds almost €7 to total costs, but interpretation of this figure is difficult in view of the nature of the research variable being used in this study.

In Table 5 we report statistics for economies of scale and scope, again for the typical university, and again using the Baumol *et al*. (1982) definitions commonly employed in this literature (for example, Johnes *et al*., 2005). This allows product-specific and ray economies of scope to be reported, where values of returns to scale in excess of unity reflect increasing returns to scale, and values below unity indicate the presence of decreasing returns to scale. A statistic is also reported for global economies of scope; a positive value for this statistic indicates that economies of scope (synergies) remain unexhausted, while a negative value suggests that there are diseconomies of scope, such that global efficiency would be enhanced if institutions were to divest some of their activities.

The findings reported in this table are striking. They suggest that there are modest ray economies of scale, but there are quite substantial diseconomies of scope. The ray scale economies come from unexhausted product-specific economies of scale for all output types. There is therefore a clear case to be made for increasing the concentration of all outputs in, and therefore the degree of specialisation of, Spanish universities.

5. Conclusions

A key advantage of using panel data to evaluate random parameter stochastic frontier models is that the attractiveness of data envelopment analysis (DEA), where each decision-making unit in effect defines its own loss function, is retained within a framework where all the tools of statistical inference are available, and where the efficiency of each unit can be evaluated. Applying this method to the case of Spanish universities serves to highlight differences in cost structures across institutions that would appear not to be due to differences in technical efficiency. Nevertheless, the question of whether the high fixed costs and the high variable costs associated with research that is observed in Pais Vasco, for example, are indeed in some sense justifiable is one that can be answered only by decision-makers. Our analysis has also served to highlight strong returns to scale and scope effects that imply that global cost savings could be realised by a reallocation of activity across universities. In themselves, these findings are sufficient to suggest that further research on this topic would be highly desirable in the Spanish context.

References

Baumol, William J., Panzar, John C. and Willig, Robert D. (1982) *Contestable markets and the theory of industry structure*, San Diego: Harcourt Brace Jovanovich.

Cohn, Elchanan Rhine, Sherrie L.W. and Santos, Maria C. (1989) Institutions of higher education as multi-product firms: economies of scale and scope, *Review of Economics and Statistics*, 71, 284-290.

Greene, William (2005) Reconsidering heterogeneity in panel data estimators of the stochastic frontier model, *Journal of Econometrics*, 126, 269-303.

Halton, John H. (1960) On the efficiency of certain quasi-random sequences of points in evaluating multidimensional integrals, *Numerische Mathematik*, 2, 84-90.

Izadi, Hooshang, Johnes, Geraint, Oskrochi, Reza, and Crouchley, Robert (2002) Stochastic frontier estimation of a CES cost function: the case of higher education in Britain, *Economics of Education Review*, 21, 63-71.

Johnes, Geraint (1996) Multi-product cost functions and the funding of tuition in UK universities, *Applied Economics Letters*, 3, 557-561.

Johnes, Geraint (1997) Costs and industrial structure in contemporary British higher education, *Economic Journal*, 107, 727-737.

Johnes, Geraint (1998) The costs of multi-product organisations and the heuristic evaluation of industrial structure, *Socio-economic Planning Sciences*, 32, 199-209.

Johnes, Geraint, Johnes, Jill, Thanassoulis, Emmanuel, Lenton, Pam and Emrouznejad, Ali (2005) *An exploratory analysis of the cost structure of higher education in England*, London: Department for Education and Skills Research Report 641.

Jondrow, James, Lovell, C.A. Knox, Materov, Ivan S. and Schmidt, Peter (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics* 19, 233-238.

Stevens, Philip A. (2005) The determinants of economic efficiency in English and Welsh universities, *Education Economics*, 13, 355-374.

Tsionas, Efthymios G. (2002) Stochastic frontier models with random coefficients, *Journal of Applied Econometrics*, 17, 127-147.

Table 1. Descriptive statistics

Variable	Mean	Standard deviation
Costs	7075.55	4246.05
Non-science undergraduates	10988.09	7050.55
Science undergraduates	9035.78	6396.09
Postgraduates	2085.67	1982.18
Research	321.14	357.81

Note: All financial variables are reported in € 0000, measured at 2000 prices.

Table 2. Results of random parameters stochastic frontier model

Variable	Coefficient
Constant	-140.755
	(0.01)
Non-science undergraduates (NS)	0.316
	(5.76)
Science undergraduates (SC)	0.196
	(4.26)
Postgraduates (PG)	0.214
	(0.84)
Research (RES)	11.023
	(8.26)
NS^2/10000	-0.194
	(5.82)
SC^2/10000	-0.062
	(1.14)
PG^2/10000	-0.266
	(0.61)
RES^2/10000	-31.757
	(1.78)
NS*SC/10000	0.261
	(4.75)
NS*PG/10000	0.402
	(2.31)
NS*RES/10000	-3.343
	(3.63)
SC*PG/10000	0.067
	(0.24)
SC*RES/10000	1.368
	(0.72)
PG*RES/10000	-2.798
	(1.16)
Standard deviation of:	
Constant	181.538
	(2.88)
RES	3.769
	(11.77)
λ	0.113
	(0.00)
σ	625.254
	(0.35)
log likelihood	-848.077

Notes: t-statistics in parentheses. For each random parameter we report the mean value of the coefficient and its standard deviation. Financial variables (costs and research) are measured in €'0000 at 2000 values. The model used is the stochastic frontier random parameters model, with normal and half-normal residuals, estimated using Limdep 8.

Table 3. Efficiencies and Slope Shifts

University	Constant	Coefficient on research	Efficiency
Almeria	-188.33	9.77	0.985
Cadiz	-155.93	10.09	0.986
Cordoba	-162.04	8.99	0.986
Huelva	-195.96	8.42	0.984
Jaen	-263.70	7.69	0.993
Oviedo	-151.49	8.09	0.991
Islas Baleares	-117.67	12.25	0.991
Las Palmas de Gran Canaria	-54.34	9.96	0.990
Castilla La Mancha	-161.92	16.80	0.991
Leon	-159.19	9.55	0.991
Salamanca	-22.24	20.53	0.991
Valladolid	-203.77	8.52	0.991
Lleida	-164.92	8.17	0.985
Politecnica de Cataluña	-170.68	11.70	0.987
Pompeu Fabra	-61.70	13.68	0.983
Rovira i Virgili	-138.61	5.94	0.983
Alicante	-111.33	14.19	0.988
Jaume I de Castellon	-170.44	8.34	0.987
Miguel Hernandez de Elche	-143.90	9.79	0.989
Santiago de Compostela	-141.41	12.90	0.988
Alcala de Henares	-197.38	9.38	0.996
Autonoma de Madrid	-110.63	11.82	0.995
Carlos III de Madrid	-105.89	8.97	0.995
Publica de Navarra	-109.56	10.62	0.995
Pais Vasco	-80.27	17.56	0.986
La Rioja	-139.31	10.95	0.987

Note: The estimated model is one in which efficiency may vary across time periods; those reported here refer to the 1998 period. The constant and research coefficient terms are invariant over the length of the panel.

Table 4. Average Incremental (AIC) costs

Output	Average Incremental Cost (€ per year)
Non-science undergraduates	3152.94
Science undergraduates	4839.52
Postgraduates	5712.36
Research	69817.48

Table 5. Economies of Scale and Scope

Ray returns to scale	1.06
Returns to scope	-0.36
Product-specific returns to:	
Non-science undergraduates	3.09
Science undergraduates	1.13
Postgraduates	1.11
Research	1.17