Lancaster University Management School
Working Paper
2005/065

Funding formulae where costs legitimately differ: the case of higher education in England

Geraint Johnes

The Department of Economics
Lancaster University Management School
Lancaster LA1 4YX
UK

FUNDING FORMULAE WHERE COSTS LEGITIMATELY DIFFER: THE CASE OF HIGHER EDUCATION IN ENGLAND

Geraint Johnes
Lancaster University Management School
Lancaster LA1 4YX
United Kingdom

T: +44 1524 594215
F: +44 1524 594244
E: G.Johnes@lancs.ac.uk

First draft: September 2005

ABSTRACT

The institutional framework for the funding of higher education in the UK is discussed. In England, much of the financial support for teaching and learning, especially of 'home and EU' undergraduates, is channelled through the Higher Education Funding Council for England (HEFCE). HEFCE operates a formula funding mechanism, though in the wake of recent policy reforms – which include the introduction of differential tuition fees – this is likely to change. Some simple economic models of funding mechanisms which may be suitable for application in this context are constructed and evaluated. Implications for the design of future policies are discussed.

JEL Classification: H42, I22
Keywords: funding, education

Paper for presentation at the December 2005 Conference on Higher Education, Multijurisdictionality and Globalisaion in Mons, Belgium.

1. Introduction

The funding of higher education in the United Kingdom has undergone dramatic change over the last few years.[1] From the perspective of students, the main changes have concerned domestic undergraduates. Some 15 years ago, mortgage-type student loans were introduced to provide partial support for maintenance (living expenses). Until then, all student maintenance was paid in the form of a grant, supplemented on the basis of a parental means test by a contribution from the student's parents. These grants were phased out over a number of years as the value of loans increased during the 1990s. Following the Dearing Report (National Committee of Inquiry into Higher Education, 1997), the mortgage loans were replaced by income-contingent loans in 1998-99, and the maintenance grant was scrapped. At the same time, a requirement was introduced that students should pay a fixed annual amount (initially £1000, but this rose steadily to £1175 in 2005-06) toward their tuition. Finally, the 2004 Education Act introduced, from academic year 2006-07, the flexibility for institutions to determine their own levels of tuition fee up to a maximum of £3000 per year. So from October 2006, tuition fees will vary both across and within institutions. This policy has been accompanied by revisions to the student loans package, and by a reintroduction, albeit at a very low level, of a means tested maintenance grant.

Undergraduate students' contributions to the costs of tuition will remain, however, only a fraction of the income generated by higher education institutions. In addition to this, institutions receive income from overseas students (who have, at institutions' discretion, paid premium fees since 1983) and all postgraduates (for whom the level of the tuition fee is likewise determined on a course by course basis by each institution). They receive income from the provision of research and outreach mission to public and private sector organisations. They also receive project-specific income from an array of seven area-specific research councils, funded by government. Finally, the Higher Education Funding Councils provide block grants to institutions to cover a portion of their teaching and research costs.

The Higher Education Funding Council for England (HEFCE)[2] provides funding for research on the basis of the numbers of research-active staff and of performance in the most recent Research Assessment Exercise. This amounts to about £1¼ billion per year. In addition, HEFCE provides over £4 billion per year in recurrent funding (and over £½ billion in capital funding) to support teaching and learning in higher education institutions. This amounts to about £2700 for each 'home and EU' student (undergraduate and postgraduate).

HEFCE's system of support for the teaching function of higher education institutions has been subject to periodic changes, but in essence the system is one in which institutions are rewarded on the basis of the numbers of student recruited. The first stage in the methodology inolves an estimation of the resource needed to deliver learning to each institution's students. This calculation is based on actual student numbers, with different weightings attached to

[1] It should be stressed that, since devolution of political powers to Wales and Scotland in 1998, differences have emerged in the funding models adopted in the different constituent countries of Great Britain. The description that follows applies to England. Note that Scotland, in particular, has followed its own course, most notably in that it does not allow its institutions to charge tuition fees for domestic undergraduates (although it does require graduates to make a financial contribution).

[2] There exist also the Higher Education Funding Council for Wales and the Scottish Higher Education Funding Council. In Northern Ireland, funding comes from the Department of Education for Northern Ireland, advised by the Northern Ireland Higher Education Council and, though a service-level agreement, by HEFCE. In this paper, our attention will be focused on the English system.

different subjects.[3] The estimated resource need (known as the 'standard resource') is then compared with a second figure known as the 'assumed resource'. This is essentially the inflation-adjusted grant that the institution received in the previous year, with allowance made for changes in fee income and other factors. So long as the standard resource is within 5 per cent of the assumed resource, the HEFCE teaching grant made to the institution in the current year equals the assumed resource. If the gap exceeds 5 per cent, HEFCE will take some action; this typically involves an adjustment in student numbers.

The ± 5 per cent tolerance band implies that institutions may vary the numbers of students that they recruit from year to year. This offers institutions, which are formally autonomous and legally independent of government, a great measure of flexibility. Premia are available to support the widening participation agenda, part-time modes of study, and location in London (where institutions face higher salary and property costs). HEFCE's methodology is therefore a formula funding method, and offers very limited scope for variation in how the formula is applied across institutions.

While this approach offers a great deal of transparency, it is a methodology that is intolerant of diversity. An institution which wished to incur high costs as a means of delivering exceptionally high quality product would find its funding squeezed as it approached the bound of the tolerance band. In the current funding model, adjustments can be made for individual institutions in some circumstances – the London weighting for example, or ad hoc arrangements for agreed additional student numbers. But these adjustments are best seen as operating outwith the funding model *per se*. This lack of flexibility, which in many respects is the key strength of any formula funding mechanism, is increasingly being viewed as a deficiency of the current model.[4]

This being so, and in light of the changes going on in other aspects of higher education funding, HEFCE is currently conducting a review of its funding mechanism for teaching and learning. A key objective of the new model is that it should more explicitly support diversity and variation in the provision of learning in higher education. In this respect at least, we can expect the new funding model to have similarities with the full Economic Costing (fEC) approach introduced in 2005 by the British research councils. This links in also to the adoption of a TRansparent Approach to Costing (TRAC), initially implemented for research costs, but currently spreading to other areas of activity within higher education institutions. The adoption of fEC in the context of research grants is widely seen to have increased income to universities. The fact that, early in 2005, HEFCE commissioned a project to explore the viability of using a cost-based approach to inform a revised teaching funding methodology suggests that the funding council is contemplating a move in the direction of fEC also for teaching.[5] It is not clear whether or not HEFCE expects institutions to gain from this, but the results obtained by Johnes *et al*. (2005) suggest that revenues for teaching currently approximately equal costs.

Nonetheless, the perception that there exists a discrepancy between 'expenditures' and 'costs' in higher education is a powerful one. This distinction is not obvious to economists who tend

[3] Subject Bands A, B and C attract 4, 1.7 and 1.3 times respectively the funding attached to Band D. Band A refers to medicine and allied subjects; Band B comprises laboratory based subjects; Band C is made up of subjects that are part classroom based and part laboratory based; and Band D subjects are classroom based.

[4] HEFCE is committed to reducing the number of 'special initiatives' that require insitutions to invest time in preparing bids for ring-fenced project funding.

[5] http://www.publications.parliament.uk/pa/cm200405/cmselect/cmsctech/220/220we119.htm

to use the terms interchangably. But HEFCE's interpretation of the terms is consistent with a view that expenditures can exceed costs owing to inefficiencies in production.[6] This means that any move toward a cost based methodology for teaching funding will likely aim to squeeze any inefficiency out of the system. Such a move is also, however, likely to mean that a greater measure of differentiation across insitutions in the remuneration offered by the funding council would be likely, this differentiation being driven by justified differences in institutions' costs. In this respect, cost based funding may be contrasted with formula funding; while, in (at least an extreme variant of) the latter case, the same formula applies to all institutions, in the former case the parameters of the formula are allowed to vary across institutions.[7] In the formal models developed later in this paper, we therefore examine both cases in which the move toward cost based funding implies a squeeze on inefficiency and cases where the nature of the funding mechanism departs from a uniform formula.

The summary given above of the funding of higher education institutions in England is necessarily a simple one. It should be clear, however, that HEFCE performs a role as a buffer organisation between government and the educational institutions. It is a quasi-non-governmental organisation (quango) with a mission to work in partnership (with institutions) to 'promote and fund high-quality, cost-effective teaching and research, meeting the diverse needs of students, the economy and society'. In a context where the development of a higher education system is planned by government, the role of a quango of this type is to free the allocation of resources from the political process. Moreover, given appropriate design of incentive mechanisms, it can do much to enhance efficiency of delivery. This type of mindset is reflected in the council's most recent strategy,[8] which states that the context within which the council operates 'implies a system that is efficiently run – one that makes good use of scarece resources. It should be a flexible, responsive and sustainable system where new needs are actively identified and met.' Implicit in this approach is the existence of twin principles – 'sustainability' (to ensure that institutions remain solvent while at least maintaining the quality of their provision) and ensuring that public funds do not subsidise private activity. This insight informs the modelling of the council's behaviour in the sequel.

Whether such a view of the role of the funding council is apposite in the context of the current system of higher education in England is moot, however. The introduction of differential tuition fees may be regarded as a simple exercise in cost sharing. In this case, HEFCE's principles as outlines above remain valid. On the other hand, a more radical interpretation of the government's approach to higher education policy would be to view differential tuition fees as a move toward a more market oriented system – one in which the disciplines of the market should apply with full force. If we take this view, the impact of HEFCE is to blunt the incentives provided by the market. By protecting institutions from financial catastrophe, the funding council could, in effect, be isolating them from the disciplines that are now supposed to be provided by the market. If the intention of the government in allowing differential tuition fees is indeed to sharpen market forces, then HEFCE as it currently operates is at odds with the thrust of the government's policy. The nature of this tension will be explored later when we introduce an analytical model of decision making under a tuition fee regime. It is a tension that remains unresolved as differential fees are about to be introduced.

[6] http://www.hefce.ac.uk/pubs/hefce/2005/05_21/05_21.pdf, especially paragraphs 64-67.

[7] In practice, funding mechanisms tend to combine elements of these two extremes. Hence, the current system includes a London weighting. It is useful therefore to think of many funding mechanisms that we observe in the world as being located along a continuum with full formula funding at one extreme and full cost based funding at the other.

[8] http://www.hefce.ac.uk/pubs/hefce/2005/05_16/05_16.pdf

In the remainder of this paper, we investigate the implications of a shift in funding council policy in the direction of a teaching funding model that can 'support diversity and variation'. While the analysis is cast in a very specialised context – that of higher education in England – it should be clear that much of what follows is, potentially at least, of much wider applicability. Indeed, while the objectives of HEFCE might be regarded as idiosyncratic, the design of reward mechanisms is an area of interest throughout the public sector and beyond. The next section provides a series of simple models that can be used to analyse the funding mechanism used by the funding body. There follows a general discussion of results and a conclusion.

2. Modelling the funding mechanism

2.1 Basic Model

To begin the formal modelling, we develop a simple model of formula funding in which the funding council allocates resources across two institutions. The institutions compete are in Nash (1951) competition with each other, and each faces a given cost technology. The objective of the funding council is to design a formula that will provide each institution with sufficient resource to break even ('sustainability'), but which will not allow either institution to make surpluses (which would enable public funding to subsidise private activity). The formula in this simple model is constrained such that average revenue is constant with respect to output. Moreover, the same formula funding mechanism applies to each institution.

Since we shall wish to use this model to investigate a transition from an expenditure based formula (where some expenditures may be inefficient) to a cost based formula, we include in the determination of expenditures an inefficiency parameter.

Consider then a game in which there are two institutions, each of which chooses its level of output so as to maximise surplus. Expenditures of the ith institution are given by

$$C_i = c + c_0 q_i + c_1 q_i^2 + \varepsilon$$

where q_i denotes the output of the ith institution and ε is the inefficiency parameter which we assume to be constant across institutions. The c, c_0 and c_1 are parameters so that c denotes the level of fixed costs that would attach to a perfectly efficient institution. Suppose further that the revenue per unit of output is determined by formula, such that revenue per unit, p, exactly matches average expenditure. Hence

$$p = \{c_0(q_1+q_2)+c_1(q_1^2+q_2^2)+2(c+\varepsilon)\}/(q_1+q_2)$$

The problem faced by each institution is therefore a Nash (1951) game with maximand

$$\max_{q_i} \{c_0(q_1+q_2)+c_1(q_1^2+q_2^2)+2(c+\varepsilon)\}q_i/(q_1+q_2) - c_0 q_i - c_1 q_i^2 - (c+\varepsilon)$$

The FOCs are given by

$$q_i^2+2q_iq_j-q_j^2 = 2(c+\varepsilon)/c_1 \qquad \forall i,\ j\neq i$$

and hence

$q_i = \sqrt{[(c+\varepsilon)/c_1]} \qquad \forall i$

Since p is set so that surplus is zero, the optimum can be thought of a point of tangency between a convex total cost curve and a linear total revenue line, the latter of which passes through the origin (see Figure 1)[9].

Note that lowering ε (as would likely happen in a move from expenditure-based to cost-based funding, for example) would shift the total cost curve down and would make the total revenue curve (which always passes through the origin) flatter, thereby reducing the level of output at which the total revenue curve is tangential to the total cost curve (Figure 2). This means that any move from expenditure-based funding to cost-based funding would be unlikely to facilitate achievement of the government's ambitious target for participation in higher education.

Some assumptions underpin the above model that might appear incongruous at first sight. First, it is assumed that institutions have the freedom to choose the level of student recruitment that they wish. The width of the ± 5 per cent tolerance band certainly gives institutions a considerable degree of discretion. But the implicit assumption here is that institutions are not constrained by demand.[10] Again, the current situation in England supports this assumption. In 2005, many prospective students failed to secure a place in any higher education institution (despite achieving the minimum requirement of two passes in the national Advanced Level examinations) owing to supply constraints.[11] Further, it is assumed that the funding council has sufficient funds to support whatever level of output the institutions choose. In reality the ± 5 per cent tolerance band protects the council from sudden changes in the claims that institutions make on its funds, but again the band is a broad one that is rarely breached in practice. Taken together, these assumptions imply that it is the technology underpinning the cost curve, and the reward structure represented by the formula funding mechanism, that determine institutions' choices about how many student places to offer.

Refinements of this model are clearly possible which would add realism. We proceed to consider, in turn: quality issues; the introduction of tuition fees; heterogeneity of costs across institution under both formula funding and differential funding; a fully flexible funding model; output target setting; and multiproduct institutions.

2.2 Quality issues

In the above model, institutions may vary q_i with impunity. There are at least two reasons why this may be unrealistic. First, the funding council may impose restrictions on the extent to which q_i may be varied from one period to the next, in the interests of stability; this is, in the HEFCE context, the case of a breach of the ± 5 per cent tolerance band. Secondly, even in the

[9] The i subscript is supressed in the diagrams.
[10] Typically institutions adjust the effective demand for their course by using grades obtained by students at the national Advanced Level examinations as a pricing mechanism.
[11] Times 30 August 2005.

absence of this type of restriction, contractual arrangements with staff may mean that it is not possible to effect a substantial reduction in supply at short notice, and hiring arrangements may likewise limit the extent to which a substantial uplift in provision is possible.

This being so, it is appropriate to consider how institutions might react if a change in the funding mechanism – a reduction in p similar to that considered earlier – were to be introduced under conditions where they are unable to change output in response. Starting from a point of tangency between the total cost curve and total revenue curve, suppose the total cost curve shifts down and the total revenue line pivots clockwise (as in Figure 2). The new tangency point would be to the left of the original solution, indicating that output should be reduced. If, however, output cannot be reduced, institutions would make losses unless they are able to reduce their costs.

But now suppose that a cost reduction might be achieved by reducing c_0 (which in all other parts of the paper we assume to be given), holding all other parameters constant. In this case, the total cost curve 'rolls up and along' the total revenue line until a new tangency point is reached at the original level of output (Figure 3).

This reduction in costs could be achieved by, for example, increasing student:staff ratios or reducing student support. This would compromise quality. So a key question to ask in considering quality issues is: how firm is the funding council's resolution to limit the extent to which institutions can vary their output from one period to the next?

The above ideas may be formalised by noting that, for *given* q_i, the level to which c_0 must fall in the ith institution in order to maintain a solution in which the gradients of the total cost and total revenue curves are equal is given by

$$c_0 = (q_1+q_2)[3q_i^2-2c_1q_i+2(c+\varepsilon)]-[c_1(q_1^2+q_2^2)+2(c+\varepsilon)]/(q_1+q_2)(1-q_i)$$

Note that the first derivative of this expression with respect to ε is unambiguously positive, indicating that quality must decline as ε falls if q_i is not allowed to vary.

Within the UK, quality is very much an issue of policy interest. Each higher education institution has procedures in place to assure and enhance the quality of its provision in the teaching and learning domain, and these are externally monitored by the Quality Assurance Agency (QAA).[12] Other mechanisms to assure quality include the external examiner system and a panoply of rankings, both based on officially collected data and on newspaper reports. In practice these mechanisms may well limit the leeway that certain institutions might have to accommodate reductions in funding by reducing quality. If it wishes to move to a 'cost based' funding system while maintaining or increasing graduate numbers, therefore, the funding council will need to be sure that institutions can (and will) bear any cuts in the unit of resource through efficiency gains, not by cutting quality.

From 2006, differential tuition fees will provide a further, and important, incentive to institutions to provide a service whose quality is demonstrably high. Differential tuition fees form the focus of discussion in our next refinement of the model.

[12] The QAA, founded in 1997, is an independent body funded by contracts with the funding councils and by subscriptions from higher education institutions.

2.3 Tuition fees

The above models assume that funding from the funding council represents the only source of income for institutions. In fact, as we have seen, there exist a plethora of other funding sources, though only a few are directly related to teaching and learning. In view of the focus of this paper on the teaching of 'home and EU' undergraduates, the introduction of differential tuition fees in 2006 is especially pertinent, and it is therefore appropriate that we should investigate the implications for our model of this policy innovation.

The models discussed above are founded on the assumption that the funding council chooses its funding mechanism to satisfy a zero-surplus constraint for each institution, assuming that each institution then chooses output to maximise surplus. If this constraint remains in place while tuition fees are introduced, no model can be tractable, since there would be no reason for institutions to prefer one level of tuition fee to another – the funding council would ensure that zero surplus is made whatever level of fee is introduced, and so institutions would be indifferent about the level of tuition fee charged.[13] The zero-surplus constraint therefore needs to be modified if the existence of differential tuition fees is to be accommodated in the model. This is done in the sequel by assuming that the funding council sets its zero-surplus constraint assuming that each institution then chooses both the tuition fee and the output level that maximises its surplus.

Suppose that each institution faces an inverse demand function given by

$$t_i = Q + \theta t_j - \rho q_i \qquad \forall i, j \neq i$$

where t_i is the tuition fee paid per student at institution i. Substituting for t_j yields

$$t_i = Q^* - \alpha q_i - \beta q_j \qquad \forall i, j \neq i$$

where Q^*, α and β are appropriately defined.

Assuming that the funding council remunerates institutions such that they make zero surplus,

$$p = \{c_0(q_1+q_2)+c_1(q_1^2+q_2^2)+2(c+\varepsilon)\}/(q_1+q_2) - Q^* + \alpha q_i + \beta q_j$$

The Nash game is then represented by the maximand facing both institutions

$$\max_{q_i} \{[c_0(q_1+q_2)+c_1(q_1^2+q_2^2)+2(c+\varepsilon)]/(q_1+q_2) - Q^* - \alpha q_i - \beta q_j\} q_i - c_0 q_i - c_1 q_i^2 - (c+\varepsilon)$$

The FOCs are given by

$$2\alpha q_i + \beta q_j + Q^* + [c_1 q_2(q_1^2 + 2q_1q_2 - q_2^2) - 2q_j(c+\varepsilon)]/(q_1+q_2)^2 = 0$$

[13] In exactly the same way, if the funding council guaranteed a zero-surplus solution in the basic model regardless of the choice of output made by institutions, no solution of the model would be possible.

Analytical solution of this model is simplified by assuming $c_1=0$. Hence, solving the pair of simultaneous equations yields

$$q_i = \{Q^* \pm \sqrt{[Q^{*2} - 2(c+\varepsilon)(2\alpha+\beta)]}\}/2(2\alpha+\beta)$$

Depending on which root represents the solution, the impact of ε on q_i may now be positive or negative. The introduction of tuition fees into the model therefore considerably complicates the analysis and removes our ability to make unambiguous statements of the impact of changes in key parameters.[14] In the remainder of the paper we therefore abstract from tuition fees in order to focus solely on the effects of funding council support.

2.4 Heterogeneity of costs and formula funding

In this section we consider the case in which costs differ across institutions and subjects. We assume that the funding council sets p so that the institution with the higher costs makes a zero surplus. In this case there is no strategic interdependence between the institutions, and solution of the model is very straightforward.

Suppose costs are given by

$$C_i = {}_ic + c_0 q_i + c_1 q_i^2 + \varepsilon$$

where ${}_ic$ denotes fixed costs of the ith institution and where ${}_2c < {}_1c$. The funding council determines p so that marginal revenue of institution 2 equals marginal cost, where the total revenue line is constrained to pass through the origin so that the solution occurs at a zero-surplus point. This is where $\partial C_i/\partial q_i = C_i/q_i$, that is where

$${}_2c + c_0 q_2 + c_1 q_2^2 + \varepsilon = c_0 q_2 + 2c_1 q_2^2$$

which implies that

$$q_2 = \sqrt{[({}_2c+\varepsilon)/c_1]}$$

and that the remuneration per unit of output, $p = c_0 + 2({}_2c+\varepsilon)/\sqrt{[({}_2c+\varepsilon)/c_1]}$.

Since the total cost curves of the two institutions are vertically parallel, $q_1=q_2$. In this case, institution 1 will make a surplus of $({}_2c - {}_1c)q_1$, and this represents a deadweight loss to the funding council.[15] This situation is illustrated in Figure 4. Of all the models considered in the present paper, this is the case that most closely matches the spirit of formula funding.

[14] Clearly if an alternative modelling strategy were adopted in which the zero-surplus constraint were not imposed, richer insights could emerge from the modelling of funding formulae in the presence of tuition fees. But such an approach would not in our view satisfactorily capture the objectives of the funding council as they presently exist.

[15] Since institutions of higher education are not-for-profit organisations, the surplus may be absorbed in the form of a transfer to research activities.

2.5 Differential funding according to full economic costs

In this section we show how the model would operate if the funding council remunerated different institutions at different levels in line with cost differentials across the institutions. Following Tinbergen (1952) it is clearly the case that the effective targeting of multiple objectives (such as the efficient determination of output in multiple institutions) requires a multiplicity of controls. Again in this case, strategic interdependence is not a feature of the model. Assume that each institution is remunerated so that it exactly covers its costs when it maximises its surplus, and – as in all previous models – the funding rule is such that zero output would entail zero remuneration. Then

$${}_ic + c_0q_i + c_1q_i^2 + \varepsilon = c_0q_i + 2c_1q_i^2 \qquad \forall i$$

so that

$$q_i = \sqrt{[({}_ic+\varepsilon)/c_1]}$$

This solution is illustrated in Figure 5. Note that this implies that higher cost institutions have higher output than lower cost institutions, and that the marginal cost is higher in the former than in the latter. This represents an allocative inefficiency, since global costs could be reduced by transferring activity from the higher cost to the lower cost institution. In the next section we consider how a funding method might be designed that finesses this problem.

2.6 A more flexible funding mechanism

Consider the case where costs are given by

$$C_i = {}_ic + c_0q_i + c_1q_i^2 + \varepsilon$$

where ${}_2c < {}_1c$. For the higher cost institution, set marginal cost equal to marginal revenue in such a manner that total revenue passes through the origin. Hence

$${}_1c + c_0q_1 + c_1q_1^2 + \varepsilon = c_0q_1 + 2c_1q_1^2$$

which implies

$$q_1 = \sqrt{[({}_1c+\varepsilon)/c_1]}$$

as before. But for institution 2, shift the total revenue function down so that it has an intercept below zero and is tangent to institution 2's total cost curve at level of output q_1. This implies that for institution 1 the revenue function is $\{c_0+2({}_1c+\varepsilon)/\sqrt{[({}_1c+\varepsilon)/c_1]}\}q_1$, but for institution 2 it is ${}_1c-{}_2c+\{c_0+2({}_1c+\varepsilon)/\sqrt{[({}_1c+\varepsilon)/c_1]}\}q_2$.

This solution is illustrated in Figure 6. Since marginal costs are now equal across institutions, this is an efficient solution. The formula adopted by the funding council in this case is one in which the per student award is constant across institutions, but a new component of the formula, related to fixed costs, is allowed to vary. This provides an efficient solution because the cost technology is also one in which fixed, but not variable, costs differ across institutions.

This would appear to be a fairly realistic scenario. The present funding model is one which has the flexibility to accommodate variation in fixed costs (owing to location in London) but not variable costs. There may of course be reasons other than location why fixed costs differ: for example some institutions have historic buildings to maintain; others have collections to insure. It would be a matter for the funding council to decide which of these differences in fixed costs are, in some sense, justifiable, and which are not. We shall say more about this later.

2.7 Demand side considerations

Hitherto, the level of output has been determined entirely by supply side considerations. In this section we consider the type of funding mechanism that might be introduced if the funding council sets (or has set for it by government) a target level, q*, for global output. Given the vertically parallel nature of the institutions' cost functions, the optimal distribution of students across the two institutions is such that $q_i=q^*/2 \ \forall i$. Following the method established earlier, we determine for each institution the equation for the total revenue curve that is tangential to the institution-specific total cost curve at output level q_i. In this case the total revenue curve is not constrained to pass through the origin for either institution.

Hence if

$$C_i = {}_ic + c_0q_i + c_1q_i^2 + \varepsilon$$

the slope of the revenue function is given by $c_0+c_1q^*$ and the intercept is ${}_ic+\varepsilon-c_1q^{*2}/4$. Hence funding from the funding council varies across institutions producing the same output, depending on the level of fixed costs, ${}_ic$. This allows the funding council to compensate institutions that face high costs owing to, say, location, while ensuring that the marginal cost of delivering output is constant across institutions so that allocative efficiency is realised. The solution is illustrated in Figure 7.

2.8 Multiproduct institutions

The above method can be generalised routinely to the case in which institutions produce multiple outputs (for instance, a multiplicity of subjects, qualifications at various levels, or study through a variety of modes of delivery) where the costs associated with each output may be distinct. It is appropriate to model such a scenario using a multiproduct cost function along the lines of those discussed by Baumol *et al*. (1982). Here we use a quadratic cost function of the form

$$C_i = {}_ic+c_0q_{1i}+c_1q_{1i}^2+d_0q_{2i}+d_1q_{2i}^2+eq_{1i}q_{2i}+\varepsilon$$

where q_{1i} represents institution i's output of type 1 and q_{2i} represents its output of type 2. Suppose that global targets are set for the production of both output types at q_1^* and q_2^*, and note that here again the need to ensure the constancy of product-specific marginal costs across institutions so that allocative efficiency obtains requires each institution to produce $q_1^*/2$ and $q_2^*/2$.

The funding council can ensure that this solution will obtain by setting the revenue function facing each institution such that the marginal revenue due to output a in each institution is equal to marginal cost, that is

$f_0+f_1q_a^*+eq_b^*/2$

where a=1,2 and b≠a, and where f_0 and f_1 denote respectively c_0 and c_1 if a=1 and d_0 and d_1 if a=2. The institution-specific intercept is then set at

${}_ic+\varepsilon-(c_1q_1^{*2}-d_1q_2^{*2}-eq_1^*q_2^*)/4$

so that each institution earns zero surplus. The introduction of multiple outputs therefore raises no new issues of principle in the model, but it should be noted that for e≠0 the interaction of outputs has an impact on the intercept.

3. Discussion and conclusions

Two important issues stand out from the above analysis. First, how can the funding model of HEFCE be reconciled with the new policy of differential tuition fees? Note was made earlier of the fact that HEFCE funding policies tend to blunt the disciplinary effects of market forces. The introduction of these market forces might detrimentally affect some institutions for reasons beyond the institutions' control – for instance some institutions may have higher fixed costs than others because their location or capital stock is in some way idiosyncratic. In such cases, financial difficulties may not be the result of managerial inefficiency, and it would be reasonable to expect HEFCE to provide additional financial support. But where an institution is in financial difficulty because of, say, mismanagement (perhaps because it has misjudged its market position and set tuition fees at an inappropriate level), it is not at all clear that HEFCE should simply fund the shortfall. In such cases, a change in the management of the institution, possibly through merger or takeover, might be a more appropriate solution. Key to HEFCE deciding on what to do in these circumstances is the availability of analysis that can distinguish between genuine cost differences between institutions on the one hand, and differences in institutional efficiency on the other.

This relates to the second issue: if institutions are to be rewarded differentially according to their costs, how can the institution-specific costs be evaluated that drive the funding model be evaluated? More specifically, how can institutions be deterred from misreporting their costs in an attempt to earn surpluses? Clearly this too is a measurement issue, where the distinction between true cost differentials and inefficiency needs to be made clear.

One possible solution lies in the estimation of statistical cost functions. The earliest modern attempt to evaluate such a function for universities is that of Cohn *et al*. (1989), though this was quickly followed by contributions by de Groot *et al*. (1991), Koshal and Koshal (1995), Glass *et al*. (1995), Johnes (1997) and others. More recent contributions have emphasised the nature of the cost curve as an envelope of efficient points in cost-output space, using stochastic frontier analysis to identify the true cost function and distinguish it from the effects of inefficiency (Johnes, 1996, 1998; Izadi *et al*., 2002; Stevens, 2005).[16] Still more recently –

[16] It could be argued that the earlier studies estimated an expenditure function rather than a cost function. The use of frontier methods fixes this problem.

and crucially in the context of the present discussion – studies have employed panel data to estimate random effect or random parameter frontier models (Johnes *et al*., 2005; Johnes and Johnes, 2005). These models have the considerable merit that they can be used to separate out genuine differences in fixed costs (and, in the random parameter case, also differences in variable costs) from differences in efficiency across higher education institutions. The econometric methods used to estimate the more sophisticated of these models are in their infancy, but the results of early analyses show great promise as a means of separating cost differentials from efficiency differentials.

The state of higher education in the UK remains one of flux. The multijurisdictionality introduced by devolved government to Wales and, particularly, Scotland in 1998 has led to differences in the ways that students are funded across parts of the United Kingdom, and the sustainability of these inter-regional variations in the face of a sharp hike in the average level of tuition fees in England and Wales remains to be seen. Likewise, it is far from clear that the existing resource allocation models used by funding councils are sustainable in a world of (relatively) high and differential tuition fees. There are clear tensions between the objectives of central government and those of the national assemblies, on the one hand, and the funding councils on the other. In the case of the national assemblies, of course, devolution has been designed to produce regional policies that are suitable for the people of the region. In the case of the funding council, no such considerations are present. It seems that the tensions between funding council policy and the policies of central government are due to a failure of the principal-agent relationship between the two. These tensions will need to be resolved.

The passage through parliament of the 2004 Higher Education Act which introduced differential fees was exciting. The Prime Minister had staked his future on the safe passage of the bill, yet one vote in the House of Commons was won by a margin of only 5 votes (even though the ruling Labour party had a majority in excess of 160 seats). The adjustments that need to be made by the assemblies and the funding councile alike will serve to ensure that the excitement is not yet over.

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 multiproduct firms: economies of scale and scope, *Review of Economics and Statistics*, 71, 284-290.

de Groot, Hans, McMahon, Walter W. and Volkwein, J. Fredericks (1991) The cost structure of American research universities, *Review of Economics and Statistics*, 73, 424-431.

Glass, J. Colin, McKillop, Donal G. and Hyndman, Noel S. (1995) Efficiency in the provision of university teaching and research: an empirical analysis of UK universities, *Journal of Applied Econometrics*, 10, 61-72.

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 (1996) 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*, Nottingham: Department for Education and Skills.

Johnes, Geraint and Johnes, Jill (2005) Higher education institutions' costs and efficiency: taking the decomposition a further step, mimeo, Lancaster University.

Koshal, Rajindar K. and Koshal, Manjulika (1995) Quality and economies of scale in higher education, *Applied Economics*, 27, 773-778.

Nash, John F. (1951) Non-cooperative games, *Annals of Mathematics*, 54, 286-295.

National Committee of Inquiry into Higher Education (1997) Higher Education in the Learning Society, Norwich: HMSO.

Stevens, Philip (2005) A stochastic frontier analysis of English and Welsh universities, *Education Economics*, forthcoming.

Tinbergen, Jan (1952) *On the theory of economic policy*, Amsterdam: North Holland.

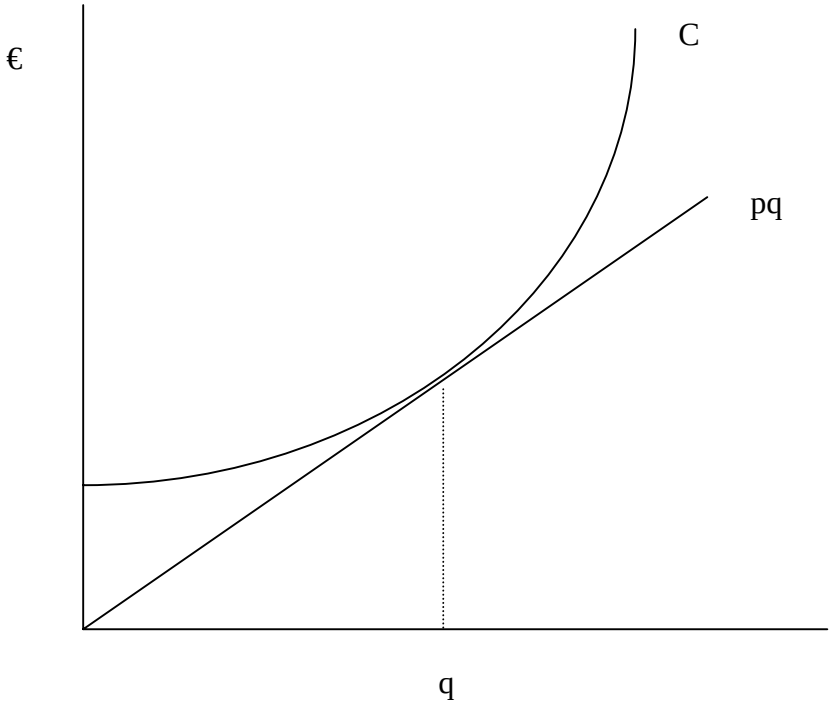

Figure 1 Basic model

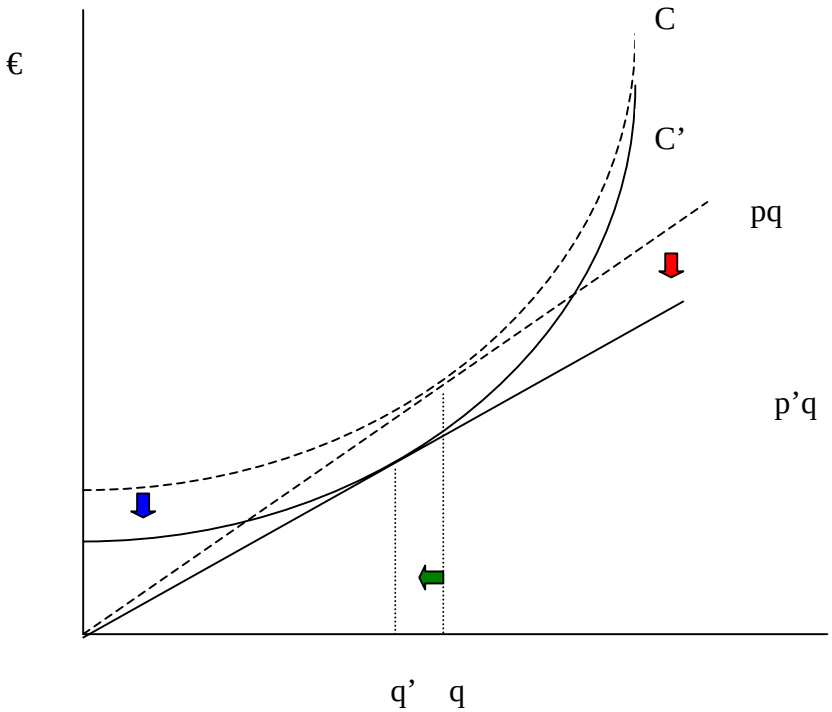

Figure 2 Effect of a cut in average revenue on output

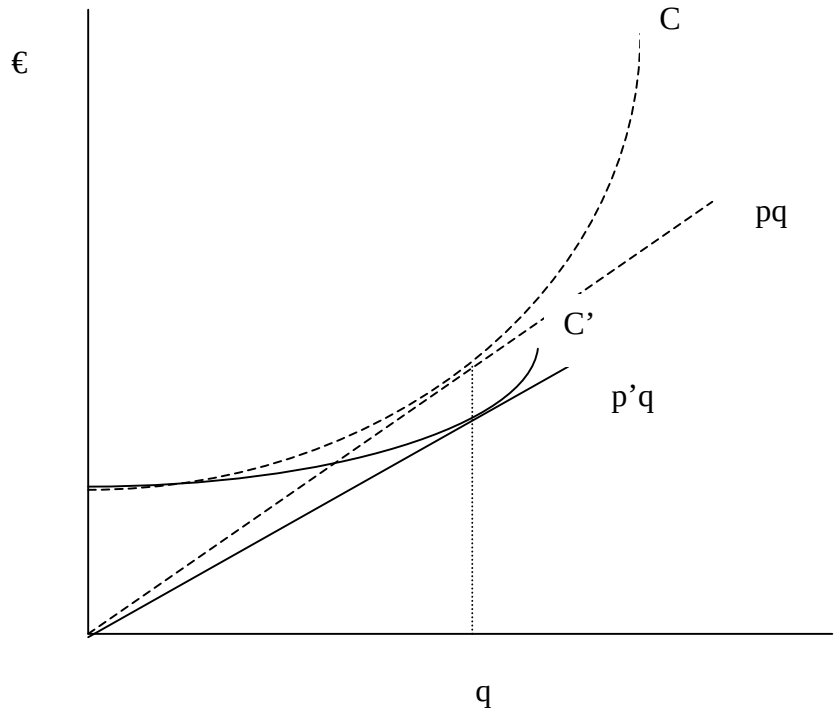

Figure 3 Imposing a quantity restriction: effects on quality

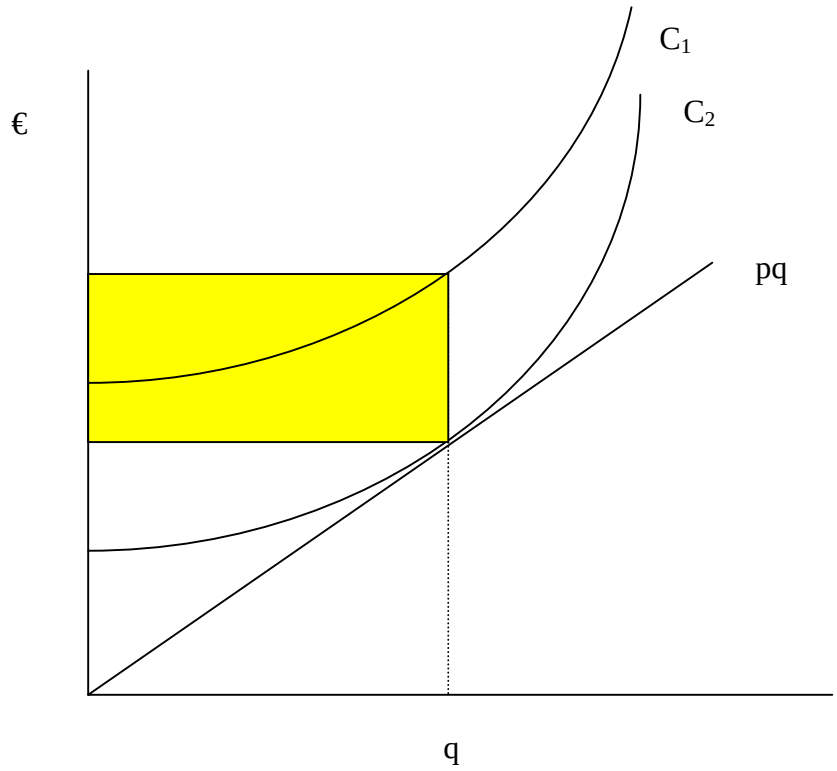

Figure 4 Deadweight associated with formula funding

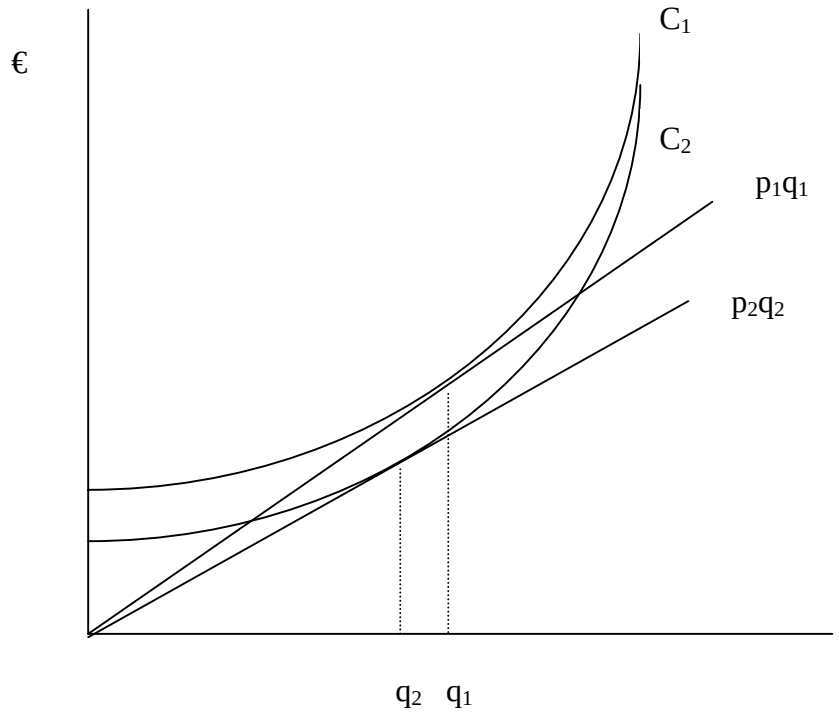

Figure 5 Cost based funding with zero intercept

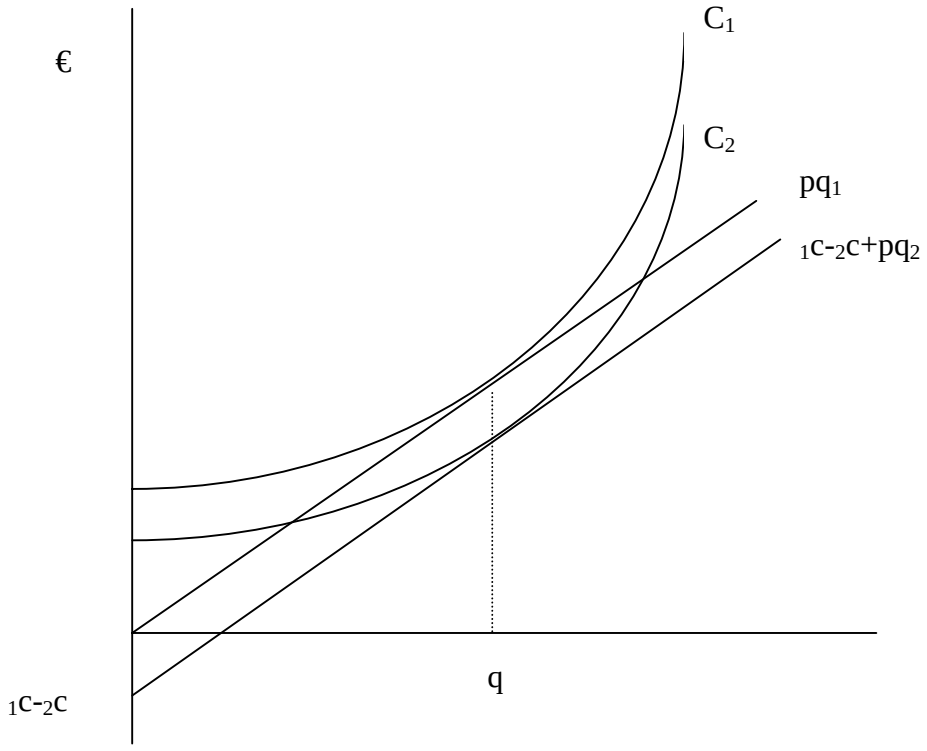

Figure 6 Cost based funding with flexible intercept

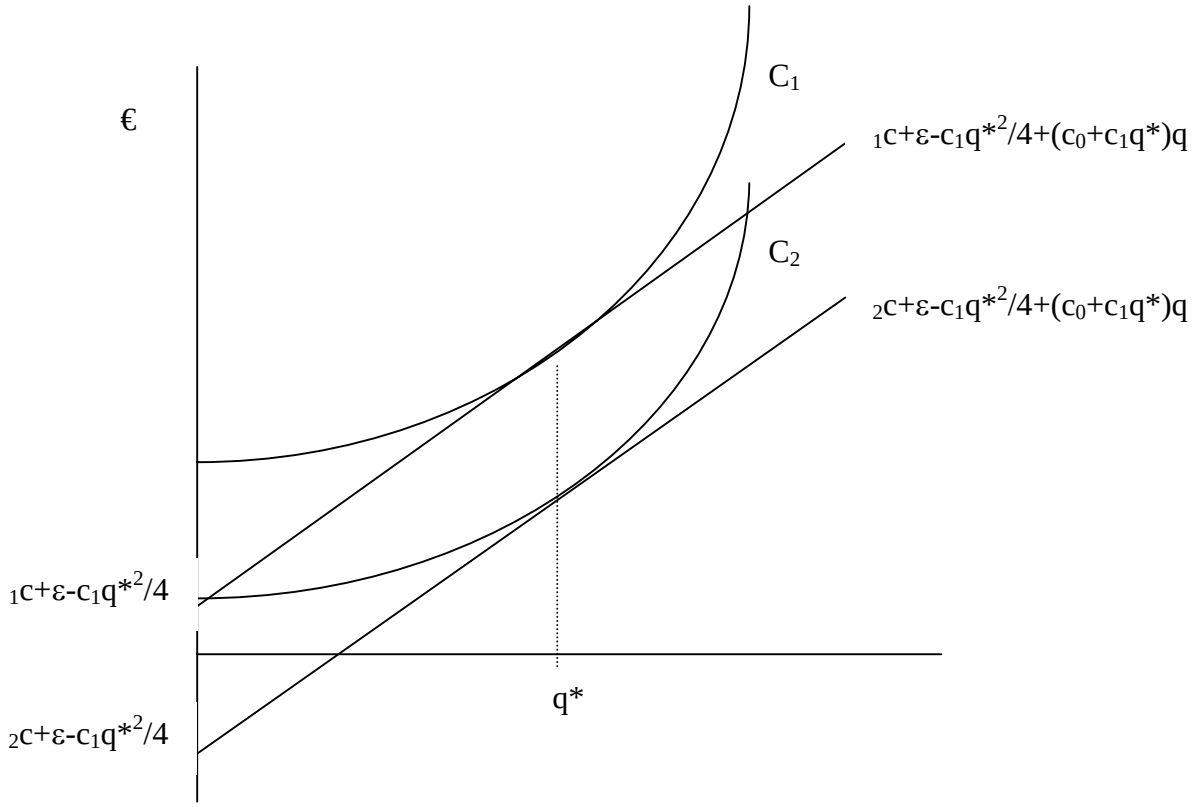

Figure 7 Cost based funding with flexible intercept: the case of fixed output