Does Expansion Cause Congestion? The Case of the Older British Universities, 1994 to 2004

A.T. FLEGG & D.O. ALLEN*

ABSTRACT *This paper examines whether the rapid growth in the number of students in British universities in recent years has led to congestion, in the sense that certain universities' output could have been higher if this expansion had been less rapid. The focus of the paper is on 45 older universities that were in existence prior to 1992. The analysis covers the period 1994/5 to 2003/4. Several alternative methods of measuring congestion are examined and, to check the sensitivity of the results to different specifications, three alternative DEA models are formulated. The results indicate that congestion was present throughout the decade under review, and in a wide range of universities, but whether it rose or fell is uncertain, as this depends on which congestion model is used. A crucial point here is whether one assumes constant or variable returns to scale. Nonetheless, all models point to a rise in congestion between 2001/2 and 2003/4, and this may well be a result of the rapid growth that occurred in this period. All models also record a sharp drop in mean technical efficiency in 2003/4. A possible explanation of the absence of a clear-cut trend in congestion is that the student : staff ratio in these universities was relatively stable in the decade under review, rising only gently from 2000/1 onwards.*

KEY WORDS: British universities; congestion; DEA

Introduction

Higher education in the United Kingdom has experienced extremely rapid growth in recent years, continuing a process that begun in the 1980s. This expansion has occurred in the 45 older universities (those existing prior to 1992), as well as in other Higher Education Institutions (HEIs). These other HEIs include the former polytechnics that became universities in 1992,

* Tony.Flegg@uwe.ac.uk and David.Allen@uwe.ac.uk. School of Economics, Bristol Business School, University of the West of England, Coldharbour Lane, Bristol BS16 1QY.

university colleges, institutes of higher education and so on. However, as illustrated in Figure 1, the growth in the number of students in the older universities since 2000/1 has been much more rapid than in the other HEIs.[1] Figure 1 also shows that, while older universities and the other HEIs have both experienced rising student : staff ratios since 2000/1, this rise has been much slower in the older universities.[2]

Figure 1 near here

The older universities and the other HEIs differ in many ways: the older universities typically have much lower student : staff ratios, substantially more research funding per member of staff, a higher proportion of undergraduate students gaining first-class degrees and upper seconds, and so on.[3] In view of these clear disparities, along with the differential rates of growth noted above, it seems appropriate to analyse the older universities and the other HEIs separately. Here we have chosen to look at the experience of the older universities in the period 1994/5 to 2003/4. This is the same group of 45 universities that we examined in our earlier study of the period 1980/1 to 1992/3 (Flegg *et al.*, 2004), thereby facilitating comparisons.

Although we believe that there are compelling arguments in favour of analysing the older universities separately from the other HEIs, other authors have taken a different view. Notable examples include Johnes (1997), Izadi *et al.* (2002) and Stevens (2005). However, none of these studies used data envelopment analysis (DEA), the technique employed here.

The Problem of Congestion

The focus of this paper is on the problem of congestion, which refers to a situation where the use of an input has increased to such an extent that output actually falls. The particular issue we wish to explore is whether the exceptionally rapid growth in the number of students in the older universities since 2000/1 has led to congestion in the sense that output could have been higher if there had been a smaller rise in the number of students.

Cooper *et al.* (2001a, p. 62) define congestion in the following way:

Definition 1. *Input congestion occurs whenever increasing one or more inputs decreases some outputs without improving other inputs or outputs. Conversely, congestion occurs when decreasing some inputs increases some outputs without worsening other inputs or outputs.*

They go on to observe (Cooper *et al.*, 2001a, p. 63) that congestion can be regarded as a particularly severe form of technical inefficiency.

However, the above definition makes no reference to any limiting factor that might account for the congestion. A possible alternative definition might read as follows:

Definition 2. *Input congestion occurs whenever more (less) of any input is employed, with all other inputs held constant, and this leads to a fall (rise) in output.*

This alternative definition is grounded in the hypothesis of diminishing marginal returns, with the added feature that congestion requires a fall (rise) in output.

Now consider the simple model $y = f(x_1, x_2)$, where y is some measure of educational output, x_1 is the number of academic staff and x_2 is the number of students. A necessary condition for congestion to exist is that one of these inputs has a negative marginal product. This will give rise to upward-sloping segments of the isoquants linking x_1 and x_2. The problem of congestion is the result of an excessive use of one or more inputs.

In the case of universities, it seems reasonable to assume that an abnormally rapid growth in the number of students could lead to congestion. For instance, Figure 1 shows that the number of full-time equivalent students in the 45 older British universities grew unusually fast from 2000/1 onwards;[4] as a result, the marginal product of students might have become *negative* in some universities. The implication of this is that a reduction in the number of students, with all other inputs (staff, buildings, etc.) held constant, would raise the university's output in terms of research and degrees awarded, both undergraduate and postgraduate. On the other hand, Figure 1 shows that this expansion was accompanied by only a very modest rise in the student : staff ratio.

Measuring Congestion

The conventional way of measuring congestion was developed by Färe and Grosskopf, while Byrnes *et al.* (1984) and Färe *et al.* (1985a) were the first published applications. Cooper *et al.* (1996) then proposed an alternative procedure, which was refined and applied to Chinese data by Brockett *et al.* (1998) and Cooper *et al.* (2000). More recently, Tone and Sahoo (2004) have proposed a new approach to measuring congestion. For ease of exposition, these alternative procedures are referred to hereafter as the approaches of Färe, Cooper and Tone.

The theoretical merits and demerits of the competing approaches of Cooper and Färe have been debated most recently by Cherchye *et al.* (2001) and Cooper *et al.* (2001a, 2001b) but no clear winner emerged from this debate. What is more, there is little published information on the extent to which these two approaches yield different outcomes in terms of the measured amount of congestion. Hence it is important to consider carefully which approach or approaches to pursue.

An important consideration is the orientation of the model. Here we would argue that an objective of maximizing output from given resources is likely to be much closer to the aims of British universities than the alternative of minimizing the resources used to produce a given output. In addition, we would maintain that the problem of congestion in British universities, if it exists, is likely to be one of excessive inputs.

However, in the current version of *OnFront*, the software supporting Färe's approach, congestion of inputs is measured using an input-oriented approach, whereas congestion of outputs is captured via an output-oriented approach. In the case of outputs, congestion refers to a situation where one or more of the outputs is an undesirable by-product of joint production, e.g. air pollution associated with the generation of electricity (cf. Färe *et al.*, 1989). Since all three outputs in our model are deemed to be desirable, congestion of outputs can be ruled out *a priori*. On the other hand, there are sound reasons for anticipating congestion with respect to one or more of the inputs.

Thus a disadvantage of using Färe's approach in the present context is that it would entail adopting an input-oriented rather than an output-oriented approach.[5] By contrast, the approaches of Cooper and Tone permit one to measure congestion of inputs via an output-oriented approach; they are, therefore, preferable in this respect. Moreover, we would argue that Färe's approach has a serious shortcoming when compared with those of Cooper and Tone: only certain instances of negative marginal productivity are deemed to constitute congestion and these cases may not even be the most plausible ones (see Flegg and Allen, 2006). Tone's approach also has the advantage that one can obtain a measure of the extent of the scale diseconomies affecting individual universities.

The most attractive feature of Cooper's approach is that it makes use of concepts that can easily be identified and measured in a set of data. What is more, his measure of congestion is easy to understand and one can immediately see which factors are apparently causing the problem and to what extent. This is more difficult to establish from Färe's procedure (see Cooper *et al.*, 2000, pp. 6–7). However, a demerit of Cooper's non-radial methodology, in comparison with Färe's radial approach, is that a straightforward decomposition of overall technical efficiency into scale, congestion and purely technical components cannot be carried out. It is also not entirely clear what aspects of the data Cooper's formula is trying to capture: is it negative marginal productivity or severe scale diseconomies or both?

In the light of the above discussion, we shall be using the approaches of Cooper and Tone as the basis for our measurements of congestion. However, later in the paper, we shall use Färe's approach to assess the sensitivity of the findings to changes in the underlying technology, i.e. to see what difference it makes if we assume constant rather than variable returns to scale.

Cooper's Approach

Cooper's approach differs in several respects from that of Färe. Färe's approach is an axiomatic one, which makes use of plausible assumptions about the nature of the productive technology (see Färe *et al.*, 1985b). It draws its inspiration from the theory of production and from the pioneering work of Farrell (1957). By contrast, Cooper's approach is more empirically based. It is grounded in the literature on data envelopment analysis (DEA).

One of the main points of contention is how input slacks should be treated. As illustrated later, an input exhibits 'slack' in situations where it is possible to reduce the quantity used of that input without causing output to decline. Färe ignores such slacks on the basis that they can be disposed of at no opportunity cost. Indeed, Färe and Grosskopf (2000a, pp. 32–33) argue that, given positive input prices, non-zero slack is akin to *allocative* rather than *technical* inefficiency. By contrast, slacks are at the core of Cooper's slacks-based measure of congestion. Cooper *et al.* (2001a, p. 69) posit a relationship of the following form:

$$c_i = s_i^{-*} - \delta_i^{*} \tag{1}$$

where c_i is the amount of congestion associated with input i, s_i^{-*} is the total amount of slack in input i and δ_i^{*} is the amount of slack attributable to technical inefficiency. The measured amount of congestion is thus a residual derived from the DEA results.

Cooper *et al.* use the following apt example to illustrate the meaning of equation (1). Consider the difference between 'an excess number of workers exhibiting idle time but not otherwise interfering with production' and 'an excess of raw material inventory congesting a factory floor in a manner that interferes with production' (Cooper *et al.*, 2001a, p. 69). The latter would represent congestion and would be captured by the variable c_i, whereas the former would represent technical inefficiency and would be measured by δ_i^{*}.

We now need to define Cooper's measure of congestion, denoted here by C_C. The first step is to rewrite equation (1) as follows:

$$c_i/x_i = s_i^{-*}/x_i - \delta_i^{*}/x_i \tag{2}$$

where c_i/x_i is the proportion of congestion in input i, s_i^{-*}/x_i is the proportion of slack in input i and δ_i^*/x_i is the proportion of technical inefficiency in input i. The second step is to take arithmetic means over all inputs to get:[6]

$$C_C = \overline{s/x} - \overline{\delta/x} \tag{3}$$

Hence C_C measures the average proportion of congestion in the inputs used by a particular decision-making unit (DMU). It has the property $0 \leq C_C \leq 1$. See Cooper *et al.* (2001a, p. 73).

The first stage of Cooper's procedure employs the Banker–Charnes–Cooper (BCC) model. In its *output-oriented* version, this involves two steps. In the first step, the model below is used to evaluate ϕ^* for each DMU k, while the second step involves maximizing the sum of the slacks, conditional on this value of ϕ^* (cf. Cooper *et al.*, 2000, pp. 3–5):

$$\phi^* = \max \phi \tag{4a}$$

subject to:

$$\textstyle\sum_j \lambda_j x_{ij} \leq x_{ik} \qquad i = 1, 2, \ldots, m \tag{4b}$$

$$\textstyle\sum_j \lambda_j y_{rj} \geq \phi y_{rk} \qquad r = 1, 2, \ldots, s \tag{4c}$$

$$\textstyle\sum_j \lambda_j = 1 \tag{4d}$$

$$\lambda_j \geq 0 \qquad j = 1, 2, \ldots, n \tag{4e}$$

To illustrate the use of Cooper's model, consider DMU E in Figure 2.[7] This diagram reveals that there are two possible referent DMUs available for evaluating E, viz B and C. Both would yield $\phi^* = 2$, yet B is the DMU that would maximize the slack in input x (giving $s_x^- = 3$ versus only 2 for C). Hence B is the DMU picked out in stage 1.

Figure 2 near here

In stage 2 of Cooper's procedure, the slacks are again maximized but subject, in this case, to the projected output remaining constant. Hence, in Figure 2, we would move along the BCC frontier from B to C, holding output constant at y = 2. This process would yield $\delta_x^* = 1$.

Thus, in the case of E, the three units of slack in input x obtained from the BCC model would be divided into two units of congestion and one unit of technical inefficiency. In terms of equation (3), we would have $\overline{s/x} = 3/5$ and $\overline{\delta/x} = 1/5$, giving $C_C = 0.4$. As regards the other DMUs, this method would generate $C_C = 0.25$ for D and F. G and H would be free from congestion, as would C. D would have $\phi^* = 2/1.5 = 1⅓$, whereas F, G and H would have $\phi^* = 2$. The figure also illustrates the point that the presence of slack is necessary but not sufficient for congestion to occur. It is worth noting, finally, that the input-oriented version of Cooper's approach would have shown no congestion for E, thereby illustrating the disadvantages of this orientation when measuring congestion of inputs (the projection would have been to point E´ in Figure 2).

In real data sets, horizontal segments such as BC in Figure 2 are rare and, in our own data set of 45 universities over 10 years, no case occurs where non-zero slack exists, yet $\phi^* = 1$. If the data set does not have any DMUs like C, then the amount of congestion for each input equals the BCC slack for this input. This greatly simplifies the work needed to compute C_C, since stage 2 of Cooper's procedure can be skipped.

A Numerical Example

To clarify the meaning of Cooper's measure, consider Figure 3.[8] This shows six hypothetical universities. Whereas university R produces an output of $y = 10$, the remaining universities all produce $y = 1$, where y is some composite index of educational output. The inputs, x_1 and x_2, represent academic staff and students, respectively. The figure takes the form of a pyramid with its pinnacle at R. R is clearly an efficient university. However, so long as *variable returns to scale* (VRS) are assumed, so too are universities A and B.[9]

Figure 3 near here

Under Cooper's approach, universities C and D would be deemed to be congested. Both are located on upward-sloping isoquant segments; this arises because $MP_1 > 0$ and $MP_2 < 0$

along segment BC, whereas $MP_1 < 0$ and $MP_2 > 0$ along segment AD. Both universities have $C_C = 0.25$, calculated as ½{(0/5) + (5/10)} for C and ½{(5/10) + (0/5)} for D. The evaluation is relative to university R in both cases.

University G is an interesting case because it is located on a downward-sloping isoquant segment; this arises because $MP_1 < 0$ *and* $MP_2 < 0$. Here C_C = ½{(2.5/7.5) + (2.5/7.5)} = ⅓. The evaluation is again relative to university R. As in the case of C and D, G is deemed to be congested because a reduction in inputs is associated with a rise in output.

However, under Färe's approach, none of these three universities would be considered to be congested! Instead, their inefficiency would be ascribed to the pure technical category. This finding can be explained by the fact that the projections onto the efficiency frontier occur along segment BA, at points C´, G´ and D´. In the identity TE ≡ PTE × SE × CE, where TE is overall technical efficiency, PTE is pure technical efficiency, SE is scale efficiency and CE is congestion efficiency, PTE = 0.4 and CE = 1 for all three universities.[10]

It is worth noting the circumstances in which a university *would* be found to be congested under Färe's approach. For instance, university C would need to be repositioned at a point such as C*, so that the ray OC* intersected the vertical line emanating from point B. Likewise, D would need to be repositioned at a point such as D*, so that the ray OD* intersected the horizontal line emanating from point A.[11] This exercise illustrates the point that an upward-sloping isoquant (negative marginal product for *one* of the factors) is necessary but not sufficient for congestion to occur under Färe's approach. In fact, for congestion to be identified, the relevant isoquant segment would need to be relatively steep or relatively flat.

University G is a rather different case: as Färe and Grosskopf (2000a, p. 32) themselves point out, a segment like CD on the unit isoquant would be ruled out of order by their axiom of *weak disposability*. In their world, isoquants may not join up in this 'circular' fashion. Weak disposability means that a proportionate rise in both x_1 and x_2 cannot reduce output.

This eliminates the possibility that both factors might have negative marginal products, which is a necessary condition for a downward-sloping segment such as CD to occur.

What might congestion mean in the case of G? Cooper *et al.* (2001a, 2001b) do not consider this issue, although they criticize Färe's approach on the grounds of its alleged adherence to the law of variable proportions. The region CDR is defined in terms of the equation $y = 28 - 1.8x_1 - 1.8x_2$, which entails that *both* marginal products must be negative. For this to make economic sense in terms of the law of variable proportions, there would need to be some latent factor that was being held constant. Alternatively, but less plausibly, one might argue that diseconomies of scale had become so severe that equiproportionate increases in both factors were causing output to fall. Cherchye *et al.* (2001, p. 77) note that this second possibility would contravene Färe's axiom of weak disposability.

From this discussion, it is clear that we should not expect the competing approaches of Cooper and Färe to yield the same outcomes in terms of congestion.[12]

A New Approach to Measuring Congestion

Tone and Sahoo (2004) have proposed a new unified approach to measuring congestion and scale economies. This has several attractive features. The first is that, unlike Färe's method, negative marginal productivity always signals congestion.[13] Secondly, the analysis can easily be done using the *DEA-Solver Pro* software (www.saitech-inc.com). Thirdly, the output is comprehensive and easily understood. For simplicity, this procedure is referred to hereafter as Tone's approach.

Tone uses an output orientation. In fact, his approach is similar to Cooper's output-oriented method inasmuch as a BCC output-oriented model is used in the first stage. However, it differs in the second stage in its use of a slacks-based measure. To explain this approach, let us return to the example in Figure 3.

Like Cooper, Tone would find A, B and R to be BCC efficient and hence not congested. The remaining DMUs would have a congestion score of $\theta = 10$, reflecting the fact that R is producing ten times as much output as any of them. A more interesting bit of output from *DEA-Solver* is the figure for the *scale diseconomy*, ρ. For example, in the case of C, this is calculated as:

$$\rho = \frac{\%\ \text{change in y}}{\%\ \text{change in } x_2} = \frac{+900\%}{-50\%} = -18 \tag{5}$$

Using the same method, we also get $\rho = -18$ for D. In the case of G, the average percentage change in inputs is $-33⅓\%$, so that $\rho = -27$. These results suggest that congestion is equally serious for C and D but more serious for G. This finding is consistent with the outcome from Cooper's approach, where $C_C = ⅓$ for G but 0.25 for C and D. In Tone's terminology, we would describe G as being *strongly* congested (because both inputs are congested) but C and D as being *weakly* congested (because only one input is congested).

Having examined the different approaches to measuring congestion, we can now consider the outputs and inputs to be used in the DEA.

Output Variables

It seems reasonable to argue that a university's output should be defined primarily according to the services it provides in terms of teaching, research, consultancy and other educational services. These aspects of a university's activities are captured here via the following output variables:

- income from research grants and contracts in £ thousands;
- the number of undergraduate degrees awarded, adjusted for quality;
- the number of postgraduate degrees, diplomas and certificates awarded.

Sources of data and other details are given in Appendix A.

Income from Research Grants and Contracts

Research is clearly an important aspect of output in its own right. It may also indirectly influence the quality of teaching output by changing the focus of a university's academic staff.

Since universities sell their services to government and industry, the income received can be used to estimate the value of the output produced. However, the use of research income as a measure of output is problematic, since such income may be held to be an input into the research process rather than an output. Research income may also be distorted by differences in research costs across academic disciplines. On the other hand, research income is likely to reflect the perceived quality, as well as quantity, of research output and it should provide a more up-to-date picture of such output than, for example, the scores in some previous research assessment exercise (cf. Stevens, 2005, p. 357). Moreover, the necessary information is readily available. Indeed, in a study of this nature, one has little option but to use research income as a proxy for research output since annual data for most alternative variables are unavailable.

Many authors have, in fact, used research income as a proxy for research output; notable examples include Johnes (1997), Izadi *et al.* (2002) and Stevens (2005). For instance, Izadi *et al.* (2002, p. 66) argue that research grants may be regarded 'as a measure of the market value of the research being undertaken [and that] their award characteristically reflects the grantee's strong research performance in the recent past.'

Income from research grants and contracts includes, *inter alia*, income received from research councils, charities, central government, local authorities, health authorities, industry, commerce and public corporations. The variable includes income from both UK and overseas sources, although income from 'other services rendered' was excluded because of concerns about the comparability of some of the data.

Undergraduate Degrees

The total number of undergraduate degrees awarded is clearly an important measure of the output of any university. However, an obvious shortcoming of this measure is that it fails to take any account of the *quality* of the degrees awarded.

One way of taking quality into account would be to multiply the number of degrees awarded by the proportion of students gaining ‘good’ degrees, defined in some way. This proportion should be positively related to the quality of teaching. With degree results, there is a choice between a very narrow definition of quality – the proportion of first-class honours degrees awarded to undergraduate students – and a broader definition comprising both firsts and upper seconds (cf. Stevens, 2005, p. 356). Here we have opted to use a broader definition of quality, so that the output variable becomes the *number* of firsts and upper seconds awarded.

Nonetheless, some problems with this output variable must be recognized. The first is that, by focusing on firsts and upper seconds, the resources employed in teaching the other graduates are being ignored. Secondly, students' achievements depend not only on the quality of teaching but also on their effort, ability and initial qualifications. Unfortunately, it was not possible to adjust for any of these attributes. A third potential problem is the possible variation, both across institutions and intertemporally, in the implicit standards set for particular classes of degree and also in the models of assessment used (e.g. the mix of coursework and examinations).

We shall be addressing the first potential problem noted above by carrying out a sensitivity analysis using *all* undergraduate awards as the output variable. As regards the third potential problem, it is worth noting that the steady long-term rise in the proportion of firsts and upper seconds awarded would not be a matter for concern, so long as it represented a genuine enhancement in the quality of graduates.[14] However, even if it did not, a common trend across all universities would not affect the DEA results.

A final caveat is worth noting: the only output recognized is degrees awarded to final-year students, despite the fact that all undergraduates are used as an input. Nonetheless, this should not present a serious problem unless the number of students was growing at substantially different rates in different universities.

Postgraduate Qualifications

For simplicity, and in order to avoid artificially boosting the efficiency scores, masters degrees and doctorates, along with postgraduate certificates and diplomas, were aggregated into a single variable.[15] A disadvantage of this is, of course, that variations across universities in the proportion of each type of postgraduate qualification are thereby ignored. This variable also fails to take account of possible differences in the quality of postgraduate qualifications.

Input Variables

The following input variables are used in the DEA analysis:

- the number of full-time equivalent undergraduate students (X_1);
- the number of full-time equivalent postgraduate students (X_2);
- academic staff expenditure in £ thousands (X_3);
- other expenditure in £ thousands (X_4).

See Appendix A for sources of data and other details. Some comments on inputs X_3 and X_4 are made below.

Academic Staff Expenditure

Input X_3 measures a university's total expenditure on academic staff. As such, it has the merit of being measured in the same units as input X_4. A possible demerit of X_3 is that staffing expenditure will vary with the proportion of staff on different grades and only approximately with the number of staff hours available for teaching, research, administration, etc. Therefore, an alternative variable – the full-time equivalent number of academic staff – is considered in a sensitivity analysis.

Other Expenditure

This variable measures a university's total expenditure *minus* its academic staff expenditure. It comprises expenditure on academic cost centres, academic services, administration and central services, premises, residences and catering, and on research grants and contracts.

Technical Efficiency

Before considering the issue of congestion, it is worth examining the overall *technical efficiency* (TE) of the 45 universities over the period 1994/5 to 2003/4.[16] Table 1 shows the results from three alternative models. *Model 1* is the one outlined above, in which the output of undergraduate qualifications is measured by the number of firsts and upper seconds awarded. In *Model 2*, this output variable is replaced by *all* undergraduate awards.[17] Finally, *Model 3* is a modified version of Model 2, whereby expenditure on academic staff is replaced by the number of full-time equivalent staff.

Table 1 near here

Along with the annual unweighted arithmetic mean (UAM) TE scores for each model, Table 1 also shows the corresponding weighted arithmetic mean (WAM) scores, which were calculated using the number of students in each university as a weight. The weighting was introduced to take account of the diverse size of universities (see Appendix B). The unweighted results, which are also illustrated in Figure 4, will be examined first.

Figure 4 near here

Table 1 and Figure 4 reveal that the unweighted mean TE scores from Model 2 typically exceed those from Model 1. This probably reflects the fact that it is possible, with Model 2, to substitute one type of undergraduate award for another, while keeping the overall number of awards constant, e.g. an upper second could be replaced by a lower second. This would tend to moderate the intertemporal fluctuations in output and lessen the variation in efficiency across universities. However, the relationship between the two graphs is not very stable.

From Table 1 and Figure 4, one can see that Models 2 and 3 yield almost identical results in the first four years. Thereafter, the graph for Model 3 lies below that for Model 2, yet it follows essentially the same pattern. This suggests that it may not make a great deal of difference to the conclusions whether one measures the input of academic staff in terms of full-time equivalents or expenditure. The close relationship between Models 2 and 3 was confirmed by the finding of a strong positive correlation (r = 0.822) between the $10 \times 45 = 450$ individual TE scores generated by each model. By contrast, r = 0.581 for Models 1 and 2.

The overall impression one gains from Figure 4 is of relatively high mean levels of technical efficiency but with no tendency for this efficiency to rise over time. Indeed, all models indicate a downturn in the final year. However, it needs to be borne in mind that the TE scores do not measure technical efficiency in an absolute sense but instead measure it relative to the frontier in each year. Hence the drop in the mean TE scores in 2003/4 could mean that the universities were moving further away from a static frontier or, alternatively, that the frontier had shifted outwards.[18] What one *can* say for sure is that there was greater *variation* in technical efficiency across universities in 2003/4 than there was in 2002/3. It is also interesting that Models 2 and 3 record a sharp drop in the number of frontier universities.

As regards the impact of weighting, the 'Difference' column in Table 1 shows a rather mixed picture. For Model 1, three of the weighted scores are higher, while seven are lower. By contrast, for Models 2 and 3, the weighting almost invariably enhances the mean scores, albeit by a modest amount in most cases. It is also worth noting that the finding of a downturn in mean TE in 2003/4 is confirmed for Models 2 and 3 but this is not so for Model 1.

Congestion: Cooper's Approach

For Cooper's approach, the mean scores were calculated by first working out C_C, the average proportion of congestion in the inputs used by each university in each year, and then averaging these figures over the 45 universities. The results are displayed in Table 2.

Table 2 near here

One can see from Table 2 that the weighted means are almost invariably less than the unweighted means. However, these differences are mostly relatively small and, for simplicity, it was decided to base the subsequent discussion on the unweighted results. Figure 5 illustrates the behaviour of the unweighted mean congestion scores obtained from the three models.

Figure 5 near here

Figure 5 reveals that Model 1 yields higher mean values of C_C than Model 2 in eight years out of ten. This divergence is especially large in the years 1995/6, 1996/7, 1997/8 and 1999/0. However, the size of the gap between the two graphs can largely be explained by differences in the extent to which each model categorizes undergraduates (input X_1) as being congested. For instance, in 1996/7, when there is a very large gap, Model 1 indicates congestion in X_1 of 6.1%, on average, whereas Model 2 yields a figure of only 1.4%, a difference of 4.7 percentage points. By contrast, there is a much smaller gap in 1998/9. Here Model 1 indicates congestion in X_1 of 4.0%, on average, whereas Model 2 yields a figure of 2.8%, a difference of only 1.2 percentage points.

The differing outcomes for Models 1 and 2 are unsurprising. This is because Model 1, by focusing on firsts and upper seconds, discriminates against universities that produce a wider range of undergraduate awards: all of their resources are counted as inputs but only part of their output is recognized, so that one might expect to see more 'congestion' as a result.

Apart from the years 1998/9, 1999/0 and 2001/2, the mean scores for Model 3 track those for Model 2 fairly closely. This is to be expected, since the two models differ only in terms of using an alternative measure of the academic staffing input. The divergence of the graphs in the three anomalous years is essentially due to different evaluations of the degree of congestion in two or three of the inputs but not, surprisingly, in the alternative academic staff variables!

It is interesting that Models 2 and 3 are in accord in suggesting that congestion rose between 2002/3 and 2003/4. This is shown by both weighted and unweighted means. Both models also find an extra six congested universities. These outcomes are consistent with the earlier finding that mean technical efficiency, as judged by these two models, fell between 2002/3 and 2003/4. However, one should bear in mind that Model 1 suggests a fall in congestion between 2002/3 and 2003/4 and also that the frontier may well have shifted between these two years.

Taking the results for the 45 universities as a whole, one might reasonably infer that congestion was not a serious problem. However, it is perhaps more meaningful to focus on the seriousness of the problem for those universities that are deemed to be congested. To illustrate, let us use Model 2 to examine the situation in 2003/4. Table 2 shows that Cooper's unweighted congestion score was 0.0233 for all universities but 0.0617 for the 17 congested universities. Having congestion equal to an average of 6.2% of inputs does seem a lot more serious.

More light can be shed on the magnitude of the problem by examining the individual values of C_C presented in Appendix B. These scores range from 0.0083 (Bath) to 0.1422 (Reading), with 12 of the 17 congested universities having $C_C > 0.040$. What we now need to do is to see how robust these findings from Cooper's approach are, by examining the results from alternative approaches. For simplicity, the discussion will be confined to Model 2.

Congestion: Tone's Approach

With Tone's approach, the following transformation was used: $C_T \equiv 1 - 1/\theta$, where $\theta \geq 1$ is the congestion score generated by *DEA-Solver Pro*. C_T can thus be compared directly with Cooper's congestion score, C_C, as both have a range from 0 (no congestion) to 1 (maximum congestion). The annual unweighted arithmetic mean values of C_T and C_C are displayed in Table 3 and illustrated in Figure 6.

Table 3 & Figure 6 near here

The results for the 45 universities show clearly that it is Tone's procedure that yields the least congestion. This is an interesting outcome because the two approaches generate exactly the same set of 17 congested universities in 2003/4; where the procedures differ is in terms of the severity of the problem in each university (see Appendix B). Here it is worth noting that not only does Tone's procedure typically indicate less congestion but it also gives a different ranking of the congested universities in most cases.

For the decade as a whole, we found only eight instances out of 450 where Tone and Cooper would disagree about whether a particular university was or was not congested (all of these had $C_T = 1$ but $C_C < 1$). This close matching of the universities held to be congested by the two approaches can be explained by the fact that both use an output-oriented version of the BCC model as their starting point. Thus scale effects are removed prior to attempting to measure congestion. Also, only those universities deemed to be inefficient in terms of the BCC model are examined for possible congestion. Therefore, even though Cooper and Tone measure congestion somewhat differently, they are still looking at the same set of potentially congested universities.

The fact that almost all congested universities have different values of C_T and C_C can be explained by the different way in which congestion is measured. For Cooper, an input exhibits congestion if it has a non-zero BCC slack, while the amount of congestion is held to be equal to that slack. The average amount of congestion over all inputs is then calculated. By contrast, Tone's procedure measures the potential increase in output from eliminating the congestion of inputs.[19] Given this difference in approach, it would be most surprising if the results *did* end up being very similar, yet Figure 6 shows that $\overline{C}_C$ and $\overline{C}_T$ do tend to follow a similar pattern, especially for the whole sample. Nonetheless, it is unclear why $\overline{C}_C$ typically exceeds $\overline{C}_T$ or, indeed, whether this is a general result.

Tone's approach also provides some useful information about scale diseconomies. Table 3 shows the annual arithmetic mean values of ρ, the scale diseconomies parameter, for the congested universities. Given a 1% decrease in congested inputs, the results indicate a potential rise in output of 3.5% on average in 2000/1 but only 1.3% in 2003/4. This suggests that congestion was more serious in 2000/1. However, given its sensitivity to extreme values,

$\bar{\rho}$ is not a very reliable measure of the amount of congestion in a given year and it is more useful to examine the values of ρ for individual universities.

To illustrate, let us consider the results for 2003/4. Appendix B shows that the value of ρ ranged from −3.544 for Hull to −0.307 for Reading. If these estimates were accurate, they would indicate that a 1% decrease in congested inputs could potentially raise output by 3.5% in Hull but by only 0.3% in Reading. It should be noted, however, that only congested inputs are included in the calculation of ρ. Likewise, only those outputs affected by congestion are considered, i.e. those where there is a potential rise in output. Hence ρ does not measure the ratio of the overall percentage changes in inputs and outputs.

Congestion: Färe's Approach

When measuring congestion, Cooper and Tone both employ an output-oriented approach, with variable returns to scale (VRS) as the underlying technology. It is, therefore, worth examining how sensitive the results are to a change in the assumed technology. Färe's approach offers a convenient way of doing this.

In their earlier work, Färe and Grosskopf assumed an absence of congestion when measuring scale effects, and only then allowed for the possibility of congestion.[20] This meant that, like Cooper and Tone, they were assuming VRS initially. However, Färe and Grosskopf (2000b) have highlighted the problems associated with distinguishing between scale inefficiency and congestion; they point out that the congestion score will depend on the *order* in which technical efficiency (TE) is decomposed. Therefore, where congestion is anticipated on *a priori* grounds, Färe and Grosskopf recommend that, rather than assuming VRS technology, one should base one's measurements on constant returns to scale (CRS). This issue will be explored here by using an input-oriented version of their approach.[21]

To clarify the relevance of the order of decomposition, consider the identity:

$$\mathrm{TE} \equiv \mathrm{PTE} \times \mathrm{SE} \times \mathrm{CE} \tag{6}$$

where PTE is pure technical efficiency, SE is scale efficiency and CE is congestion efficiency. Crucially, in this identity, TE and the product SE × CE are unaffected by the order of the decomposition but the individual values of SE and CE *are* affected.

A glance at Figure 7 is all that is needed to see that we get far more 'congestion' if we assume CRS rather than VRS. What is more, the gap between the $\overline{C}_{F,CRS}$ and $\overline{C}_{F,VRS}$ graphs is much wider at the end of the period than it is at the beginning. Cooper's measure, $\overline{C}_{C}$, also stands out as being the most different from the other two VRS-based measures. Another important finding is that there is hardly any difference between Färe's VRS-based measure and that of Tone. These findings are substantiated in Table 4. It may be noted, finally, that all four measures show a rise in mean congestion in the last year.

Tables 4 & 5 near here

To shed some more light on the relationships among the different measures, correlation coefficients were calculated using the raw congestion scores (n = 450). Table 5 shows the results. As expected, Färe's VRS-based measure is very strongly correlated with that of Tone. The fact that this correlation is 0.944 rather than unity can be attributed to two factors: the different orientation and the different ways in which congestion is measured. Here it is worth noting that Färe uses a radial (i.e. proportional) projection to eliminate congestion, whereas Tone uses a slacks-based measure. However, neither the different orientation nor the different projection employed appears to make a great deal of difference to the results. What is of most importance is whether one uses CRS or VRS.[22]

Cooper's measure is, as anticipated, more strongly correlated with $C_{F,\,VRS}$ than it is with $C_{F,\,CRS}$. Even so, C_C and $C_{F,\,VRS}$ are clearly not very close substitutes. Taken as a whole, the correlation analysis shows that the four measures tend to move in the same direction. However, the strength of this correlation varies substantially across the four measures and some are clearly more substitutable than others. More detailed information is given in Appendix B, where the individual results for 2003/4 are tabulated.

Decomposing Congestion

A helpful feature of Cooper's procedure is that it allows one to measure, for each congested university, how much each input contributes to the observed amount of congestion. Table 6 takes a closer look at this facet of his approach, using annual means to summarize the data. The table shows the contribution of each input to the annual unweighted mean value of C_C.

Table 6 near here

The results for Model 1 reveal that excessive numbers of students − both undergraduate and postgraduate − were the predominant cause of congestion in British universities during the decade under review. Students accounted for 75.2%, on average, of the value of $\overline{C}_C$, whereas academic staff accounted for 18.7% and 'other expenditure' for a mere 6.1%.

Rather different results are obtained from Model 2. There is a marked fall in the proportion of congestion attributable to undergraduates and a concomitant rise in that due to postgraduates. However, at 72.3%, the combined share of these two inputs is only marginally lower. The fall in undergraduates' share can probably be explained by the fact that Model 2 recognizes a wider range of undergraduate awards as being relevant, so that fewer undergraduates are deemed to be redundant. A similar explanation can be adduced to account for the fall in the share of academic staff. There is also a noticeable rise in the share of 'other expenditure', although this is still only 12.5%.

Models 2 and 3 produce fairly similar results, as might be expected. Even so, some changes are worth noting. The first is that, as a result of using full-time equivalents rather than expenditure, congestion due to academic overstaffing has fallen by 2.7 percentage points. Secondly, the combined share of students has fallen from 72.3% to 67.1%. Finally, there is a further rise, from 12.5% to 20.4%, in the share of 'other expenditure'.

Whilst it is easy to understand why having too many students could lead to congestion, it is more challenging to explain why academic staff might be congesting. One possible explanation is that overstaffing caused congestion of facilities such as libraries, office accommodation, etc. and this, in turn, caused a fall in output. This could be relevant if the frontier universities were generally better endowed than the congested universities. It is also possible that the 'surplus' staff in the congested universities might be indicative of a more general form of institutional inefficiency.

The role of 'other expenditure' in generating congestion is also puzzling. What the results suggest is that, beyond a certain point, extra expenditure actually reduced congested universities' output. However, a possible explanation is in terms of the *mix* of expenditure. 'Other expenditure' is a very broadly defined input variable, comprising expenditure on academic cost centres, academic services, administration and central services, premises, residences and catering, and on research grants and contracts. It is conceivable that a rise in the proportion of other expenditure devoted to research could lower the output of undergraduate and postgraduate qualifications, even though it might stimulate research activity. Conversely, a fall in this proportion could have the opposite effect. Another possible explanation is in terms of excessive spending on administration, which could reduce a university's efficiency and hence output in terms of research and qualifications awarded.

Conclusion

This paper has examined the performance of 45 British universities that were in existence prior to 1992, using annual data for the period 1994/5 to 2003/4. This decade witnessed rapid growth in the number of full-time equivalent students in these older universities: this number grew from 560,000 in 1994/5 to 762,000 in 2003/4, a rise of 36%. The growth was particularly fast in the last three years, which accounted for 94,000 of the 202,000 extra students. An interesting question is whether this exceptionally rapid growth caused congestion, in the sense that certain universities' output could have been higher if the expansion had been less rapid.

Three alternative approaches to measuring congestion were examined: the conventional approach of Färe and Grosskopf, the alternative proposed by Cooper *et al.*, and a new method developed by Tone and Sahoo. In addition, in the case of Färe and Grosskopf's approach, two versions were considered: one assumed constant returns to scale (CRS), while the other assumed variable returns to scale (VRS). To check the sensitivity of the results to different specifications, three alternative DEA models were formulated.

Tone and Sahoo's method and the VRS-based version of Färe and Grosskopf's approach generated almost identical results; these results indicated a *fall* in average congestion scores over the decade as a whole but a modest rise in the last two years. Cooper's method typically produced noticeably higher congestion scores than the other two methods, especially from 1998/9 onwards. This method suggested a slight fall in congestion over the decade as a whole, yet a marked rise in the last two years.

Switching from VRS to CRS had a dramatic impact on the results generated by Färe and Grosskopf's approach: the mean congestion scores were much higher in almost all years and there was a particularly sharp rise in the final year. In contrast to the other methods, all of which are based on VRS, this CRS-based method suggested a marked *rise* in congestion over the decade as a whole, from an average of 3.3% of inputs in 1994/5 to 4.6% in 2003/4.

Thus the evidence on whether 'expansion causes congestion' is rather mixed. Certainly, congestion was present throughout the decade under review, and in a wide range of universities, but whether it increased or decreased depends on which model one looks at! Nonetheless, all models pointed to a rise in congestion between 2001/2 and 2003/4, and this may well be a consequence of the rapid expansion that occurred over this period. It is worth noting too that all models indicated a sharp drop in mean technical efficiency in 2003/4.

A possible explanation of the absence of a clear-cut trend in congestion is that the student : staff ratio in the 45 older British universities was relatively stable during the decade under review, rising only gently from 2000/1 onwards. This indicates that the number of academic staff also grew rapidly over the decade, albeit not as fast as the number of students.

When the congestion scores from the various models were averaged over all universities, congestion did not appear to be a particularly serious problem. However, it was argued that it was more realistic to include only those universities that were found to be congested. Doing this had the effect of raising the means substantially. For example, in the case of Cooper's method, the unweighted mean congestion score for 2003/4, based on our DEA Model 2, rose from 0.0233 to 0.0617 when the scores were averaged over the 17 congested universities rather than over all 45 universities. When looked at in this way, with congestion averaging 6.2% of inputs, the problem seemed rather more serious.

A decomposition analysis using Cooper's procedure revealed that an overabundance of students − both undergraduate and postgraduate − was the primary cause of congestion throughout the period under review. For instance, in 2003/4, our DEA Model 2 shows that students accounted for over 77% of the average value of Cooper's measure of congestion. Academic staff and 'other expenditure', by contrast, had a far smaller role in generating congestion.

There are some areas where this study could usefully be built upon. The first is that a Malmquist analysis could be employed to distinguish between fluctuations in congestion brought about by shifts in the efficiency frontier, as opposed to movements towards or away from this frontier. Secondly, use could be made of the facility in *OnFront*, whereby one can restrict consideration to a subset of inputs most likely to be affected by congestion. Thirdly, it would be useful to explore why Cooper's approach typically generated higher congestion scores than the other VRS-based methods. Finally, it would be interesting to see whether a university's discipline mix has an effect on its congestion scores.

Acknowledgements

We would particularly like to thank Timo Kuosmanen, and the anonymous referees, for some very helpful suggestions. We have also benefited from discussions with John Sloman and Chris Webber. Finally, we would like to thank Kate Lang of the Higher Education Statistics Agency (HESA) for the efficient way in which she answered our many queries and produced the data we needed. See Appendix A for details.

Notes

1. The basic data used in this study were obtained, either directly or indirectly, from the Higher Education Statistics Agency (HESA). See Appendix A for details.

2. The ratios have not been plotted for the first two years because of concerns about the quality of the data for full-time equivalent staff and students. See Appendix A.

3. In 2003/4, for example, 13.2% of undergraduates in the 45 older universities gained first-class degrees and 49.3% gained upper seconds, whereas the proportions in the other HEIs were 8.0% and 39.3%, respectively. *Source*: Authors' own calculations using HESA data.

4. The annual rates of growth were 4.7%, 4.9% and 3.9%, respectively, in the last three years.

5. Pontus Roos, of the Institute of Applied Economics in Sweden, has told us that it will be possible, in a forthcoming new version of *OnFront*, to measure congestion of inputs via an output-oriented approach. For details of the *OnFront* software, see www.emq.com.

6. There is a case for using geometric rather than arithmetic means to average these ratios.

7. Figure 2 is adapted from Brockett *et al.* (1998).

8. Figure 3 is adapted from Cooper *et al.* (2001a).

9. A and B would be inefficient under constant returns to scale whereas R would be efficient.

10. This was confirmed using *OnFront* and an input-oriented model. TE = 0.1 and SE = 0.25 for C and D; G has TE = $0.0\dot{6}$ and SE = $0.1\dot{6}$.

11. CE = Oc/OC* and CE = Od/OD* for the repositioned C and D, where CE ≈ 0.6 in both cases.

12. For a more detailed discussion, see Flegg and Allen (2006).

13. We are indebted to Kaoru Tone for confirming this point.

14. In 2003/4, for example, 13.2% of undergraduates in the 45 older universities gained first-class degrees and 49.3% gained upper seconds, compared with 9.6% and 45.5%, respectively, in 1994/5. *Source*: Authors' own calculations using HESA data.

15. DEA efficiency scores tend to rise as the number of variables increases, thereby reducing the discriminatory power of the technique.

16. These TE scores were obtained (using *DEA-Solver-Pro*) from the Charnes–Cooper–Rhodes (CCR) model, which assumes constant returns to scale (CRS) and no congestion. The orientation of the model has no effect on the TE scores under CRS. *OnFront* generated identical results.
17. This broader variable encompasses all undergraduate degrees, as well as 'other undergraduate awards' such as certificates and diplomas in business, computing, engineering, medicine, nursing and technology, along with higher national diplomas, certificates and diplomas of higher education, etc. For the 45 older universities, these 'other awards' increased in importance from 4.5% of all undergraduate awards in 1994/5 to 10.7% in 2003/4. In some cases, these other awards are a default qualification rather than one that would be sought in its own right.
18. To discriminate between these two possibilities would require a Malmquist analysis (see Flegg *et al.*, 2004). However, an analysis of this kind is beyond the scope of the present paper.
19. Tone uses an output-oriented slacks-based measure in his projection of the congested universities onto the BCC frontier. For an explanation of this SBM procedure, see Tone (2001).
20. See, for example, Byrnes *et al.* (1984), and Färe *et al.* (1985a).
21. The calculations were carried out using *OnFront*. For comparative purposes, congestion efficiency (CE) scores were converted into *in*efficiency scores by defining $C_F \equiv 1 - CE$.
22. In our earlier study, Flegg *et al.* (2004), we used an output-oriented variant of Färe's approach and assumed VRS. When we reworked the results using Tone's approach, we got remarkably similar congestion scores.

Appendix A. Sources and definitions

Most of the data used in this study were obtained directly from various issues of the following publications of the Higher Education Statistics Agency (HESA):

- *Resources of Higher Education Institutions* (RHEI)
- *Students in Higher Education Institutions* (SHEI)

See HESA (various years). In some cases, noted below, data were obtained directly from HESA under contract. The results for the first two years of our study need to be treated cautiously, owing to possible problems with the data on full-time equivalent numbers of students and staff.

For simplicity, and in order to facilitate comparisons with our earlier study, data for the constituent colleges and institutes of the University of London were aggregated. Likewise, data for the member institutions of the University of Wales were aggregated, although we did not include data for associated institutions (e.g. Swansea Institute of Higher Education). University of Wales, Cardiff (later classified as Cardiff University) was included throughout.

Some key information on the variables used in this study is given below. More detailed information is given in the HESA publications mentioned above.

- *Income from research grants and contracts*

Because of concerns about the comparability of some of the data, especially for Cambridge, this variable excludes data on what HESA defines as income from 'other services rendered'. Source: RHEI, Table 3 up to 2001/2, Table 1c thereafter.

- *Number of undergraduate and postgraduate qualifications awarded*

The qualifications data published in SHEI could not be used for two reasons:

(i) the severe rounding of the published data in the last five years;

(ii) the unspecified qualifications of 'dormant students' in the first six years.

Fortunately, we were able to obtain the necessary data directly from HESA.

- *Full-time equivalent undergraduate and postgraduate students* (X_1 and X_2)

HESA did not publish full-time equivalent numbers for 1994/5 and 1995/6, owing to concerns about the quality of the data. Whilst we were able to obtain the unpublished data directly from HESA, we have used these figures in our study with some reservations. Data from 1996/7 onwards were obtained from SHEI, Table 0b.

- *Academic staff expenditure* (X_3)

Source: RHEI, Table 7 up to 2001/2, Table 2b thereafter.

- *Other expenditure* (X_4)

Variable X_4 was calculated by subtracting what HESA defines as 'other expenditure' from each university's total expenditure and then deducting academic staff expenditure (X_3). HESA's 'other expenditure' was not included, as we were concerned about the comparability of some of the data.

Source: RHEI, Tables 6 and 7 up to 2001/2, Tables 2a and 2b thereafter.

- *Full-time equivalent number of academic staff*

The HESA data on this variable were downloaded from http://www.data-archive.ac.uk. It should be noted that we have some concerns about the reliability of the data for 1995/6. In particular, the aggregate student : staff ratio for that year looks unrealistically high.

Appendix B. Individual results for 2003/4: Model 2

				Färe				**Tone**			**Cooper**	
	weight	**TE**	**rank**	**$C_{F, CRS}$**	**rank**	**$C_{F, VRS}$**	**rank**	**C_T**	**rank**	**ρ**	**C_C**	**rank**
Aston	0.009	0.8494	37	0.1510	42	0	1	0	1		0	1
Bath	0.013	0.8647	33	0.0980	38	0.0163	7	0.0280	38	–2.094	0.0083	29
Birmingham	0.031	0.9472	23	0.0530	33	0	1	0	1		0	1
Bradford	0.011	0.8307	41	0.1690	43	0	1	0	1		0	1
Bristol	0.020	0.8917	28	0.0200	20	0.0165	8	0.0153	33	–1.502	0.0677	39
Brunel	0.016	0.9206	26	0.0790	35	0.0230	13	0.0282	39	–1.993	0.1250	44
Cambridge	0.024	1	1	0	1	0	1	0	1		0	1
City	0.017	1	1	0	1	0	1	0	1		0	1
Durham	0.018	0.9563	22	0.0440	29	0.0360	15	0.0324	41	–2.209	0.0693	40
East Anglia	0.013	0.9765	17	0.0230	22	0	1	0	1		0	1
Essex	0.010	0.8357	39	0.0470	31	0	1	0	1		0	1
Exeter	0.014	0.8448	38	0.0470	31	0.0230	12	0.0379	43	–1.201	0.0633	38
Hull	0.017	0.8873	30	0.1130	40	0.0710	17	0.0838	44	–3.544	0.0587	36
Keele	0.010	0.8495	36	0.0150	18	0	1	0	1		0	1
Kent	0.015	0.9363	25	0.0640	34	0	1	0	1		0	1
Lancaster	0.014	1	1	0	1	0	1	0	1		0	1
Leeds	0.037	0.9713	19	0.0290	26	0	1	0	1		0	1
Leicester	0.017	0.9798	15	0.0200	20	0.0190	10	0.0198	37	–0.343	0.0614	37
Liverpool	0.023	1	1	0	1	0	1	0	1		0	1
London	0.127	1	1	0	1	0	1	0	1		0	1
Loughborough	0.017	0.9471	24	0.0360	28	0.0161	6	0.0187	35	–1.316	0.0327	32
Manchester	0.033	0.8516	35	0.0020	14	0.0165	9	0.0159	34	–0.807	0.0179	31
UMIST	0.009	0.8242	42	0.1760	44	0	1	0	1		0	1
Newcastle	0.022	0.8909	29	0.0070	15	0.0067	4	0.0132	32	–1.314	0.0477	35
Nottingham	0.032	0.9813	14	0.0190	19	0	1	0	1		0	1
Oxford	0.024	1	1	0	1	0	1	0	1		0	1
Reading	0.015	0.9775	16	0.0230	22	0.0060	2	0.0086	29	–0.307	0.1422	45
Salford	0.022	1	1	0	1	0	1	0	1		0	1
Sheffield	0.029	0.9732	18	0.0270	25	0	1	0	1		0	1
Southampton	0.024	0.9872	13	0.0130	17	0.0110	5	0.0113	31	–0.679	0.0879	42
Surrey	0.013	0.8348	40	0.0010	13	0.0032	1	0	1		0	1
Sussex	0.012	0.8820	31	0.0440	29	0.0060	2	0.0096	30	–0.467	0.0146	30
Warwick	0.020	1	1	0	1	0	1	0	1		0	1
York	0.012	1	1	0	1	0	1	0	1		0	1
Wales	0.080	0.8214	43	0.0080	16	0	1	0	1		0	1
Aberdeen	0.015	1	1	0	1	0	1	0	1		0	1
Dundee	0.017	1	1	0	1	0	1	0	1		0	1
Edinburgh	0.025	1	1	0	1	0	1	0	1		0	1
Glasgow	0.024	0.9701	20	0.0300	27	0.0300	14	0.0284	40	–0.395	0.0827	41
Heriot-Watt	0.009	0.7471	45	0.2530	45	0	1	0	1		0	1
St. Andrews	0.010	0.7679	44	0.0850	36	0	1	0	1		0	1
Stirling	0.009	0.8637	34	0.1360	41	0	1	0	1		0	1
Strathclyde	0.021	0.8741	32	0.1020	39	0.1200	18	0.0995	45	–1.524	0.0395	33
Belfast	0.022	0.9686	21	0.0260	24	0.0210	11	0.0191	36	–1.240	0.0413	34
Ulster	0.027	0.9077	27	0.0920	37	0.0490	16	0.0327	42	–1.800	0.0884	43
Mean		0.9247		0.0456		0.0109		0.0112		–1.337	0.0233	
Number on frontier		12		12		27		28			28	
Correlations TE				–0.680		–0.128		–0.152			–0.061	
$C_{F, CRS}$						0.208		0.215			0.004	
$C_{F, VRS}$								0.969			0.475	

References

Brockett, P.L., Cooper, W.W., Shin, H.C. & Wang, Y. (1998) Inefficiency and congestion in Chinese production before and after the 1978 economic reforms, *Socio-Economic Planning Sciences*, 32, pp. 1–20.

Byrnes, P., Färe, R. & Grosskopf, S. (1984) Measuring productive efficiency: an application to Illinois strip mines, *Management Science*, 30, pp. 671–681.

Cherchye, L., Kuosmanen, T. & Post, T. (2001) Alternative treatments of congestion in DEA: a rejoinder to Cooper, Gu, and Li, *European Journal of Operational Research*, 132, pp. 75–80.

Cooper, W.W., Gu, B. & Li, S. (2001a) Comparisons and evaluations of alternative approaches to the treatment of congestion in DEA, *European Journal of Operational Research*, 132, pp. 62–74.

Cooper, W.W., Gu, B. & Li, S. (2001b) Note: alternative treatments of congestion in DEA – a response to the Cherchye, Kuosmanen and Post critique, *European Journal of Operational Research*, 132, pp. 81–87.

Cooper, W.W., Seiford, L.M. & Zhu, J. (2000) A unified additive model approach for evaluating inefficiency and congestion with associated measures in DEA, *Socio-Economic Planning Sciences*, 34, pp. 1–25.

Cooper, W.W., Thompson, R.G. & Thrall, R.M. (1996) Introduction: extensions and new developments in DEA, *Annals of Operations Research*, 66, pp. 3–45.

Färe, R. & Grosskopf, S. (2000a) Slacks and congestion: a comment, *Socio-Economic Planning Sciences*, 34, pp. 27–33.

Färe, R. & Grosskopf, S. (2000b) Research note: decomposing technical efficiency with care, *Management Science*, 46, pp. 167–168.

Färe, R., Grosskopf, S. & Logan, J. (1985a) The relative performance of publicly-owned and privately-owned electric utilities, *Journal of Public Economics*, 26, pp. 89–106.

Färe, R., Grosskopf, S. & Lovell, C.A.K. (1985b) *The Measurement of Efficiency of Production* (Boston, Kluwer-Nijhoff).

Färe, R., Grosskopf, S., Lovell, C.A.K. & Pasurka, C. (1989) Multilateral productivity comparisons when some outputs are undesirable: a non-parametric approach, *Review of Economics and Statistics*, 71, pp. 90–98.

Farrell, M.J. (1957) The measurement of productive efficiency, *Journal of the Royal Statistical Society*, Series A, General, 120, Part 3, pp. 253–281.

Flegg, A.T., Allen, D.O., Field, K. & Thurlow, T.W. (2004) Measuring the efficiency of British universities: a multi-period data envelopment analysis, *Education Economics*, 12, pp. 231–249.

Flegg, A.T. & Allen, D.O. (2006) An examination of alternative approaches to measuring congestion in British universities, Discussion Paper No. 06/06, School of Economics, Bristol Business School, University of the West of England, Bristol (http://carecon.org.uk/DPs/).

HESA (various years) *Resources of Higher Education Institutions* and *Students in Higher Education Institutions* (Cheltenham, Higher Education Statistics Agency).

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

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

Stevens, P.A. (2005) A stochastic frontier analysis of English and Welsh universities, *Education Economics*, 13, pp. 355–374.

Tone, K. (2001) A slacks-based measure of efficiency in data envelopment analysis, *European Journal of Operational Research*, 130, pp. 498–509.

Tone, K. & Sahoo, B.K. (2004) Degree of scale economies and congestion: a unified DEA approach, *European Journal of Operational Research*, 158, pp. 755–772.

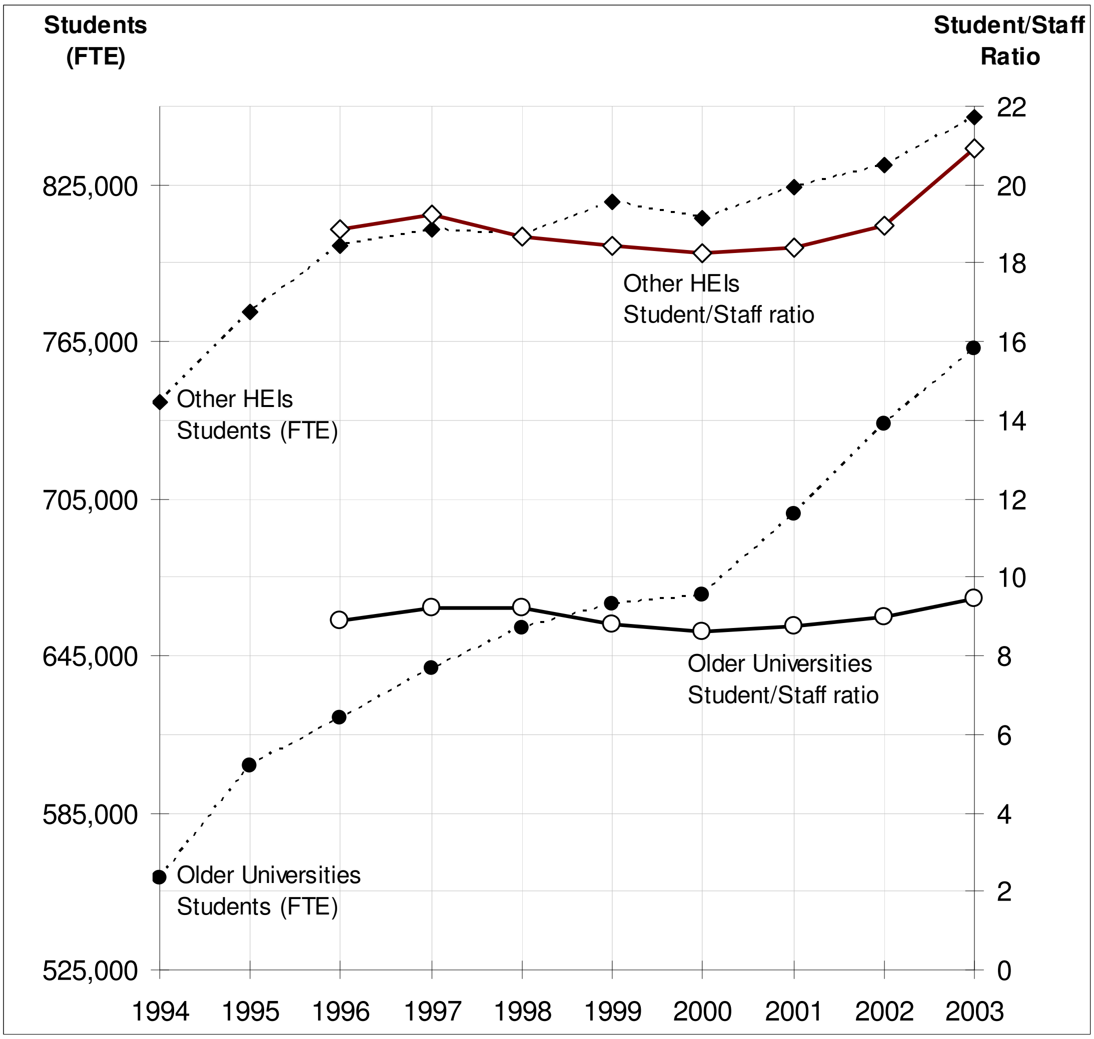

Fig. 1. Comparison of older universities with other HEIs

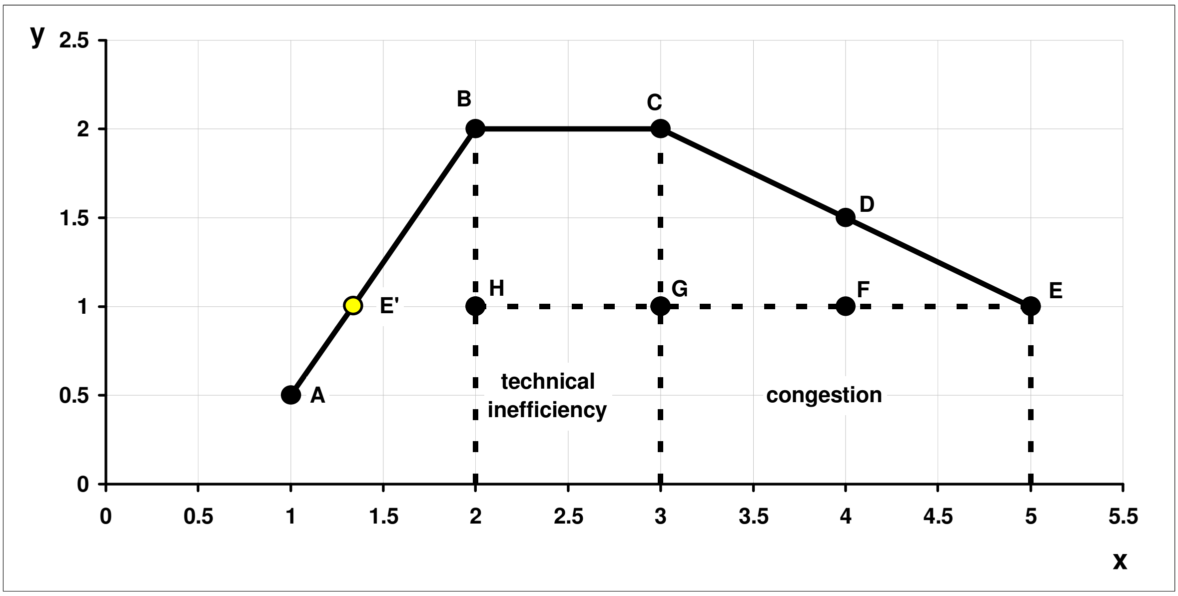

Fig. 2. Cooper's model

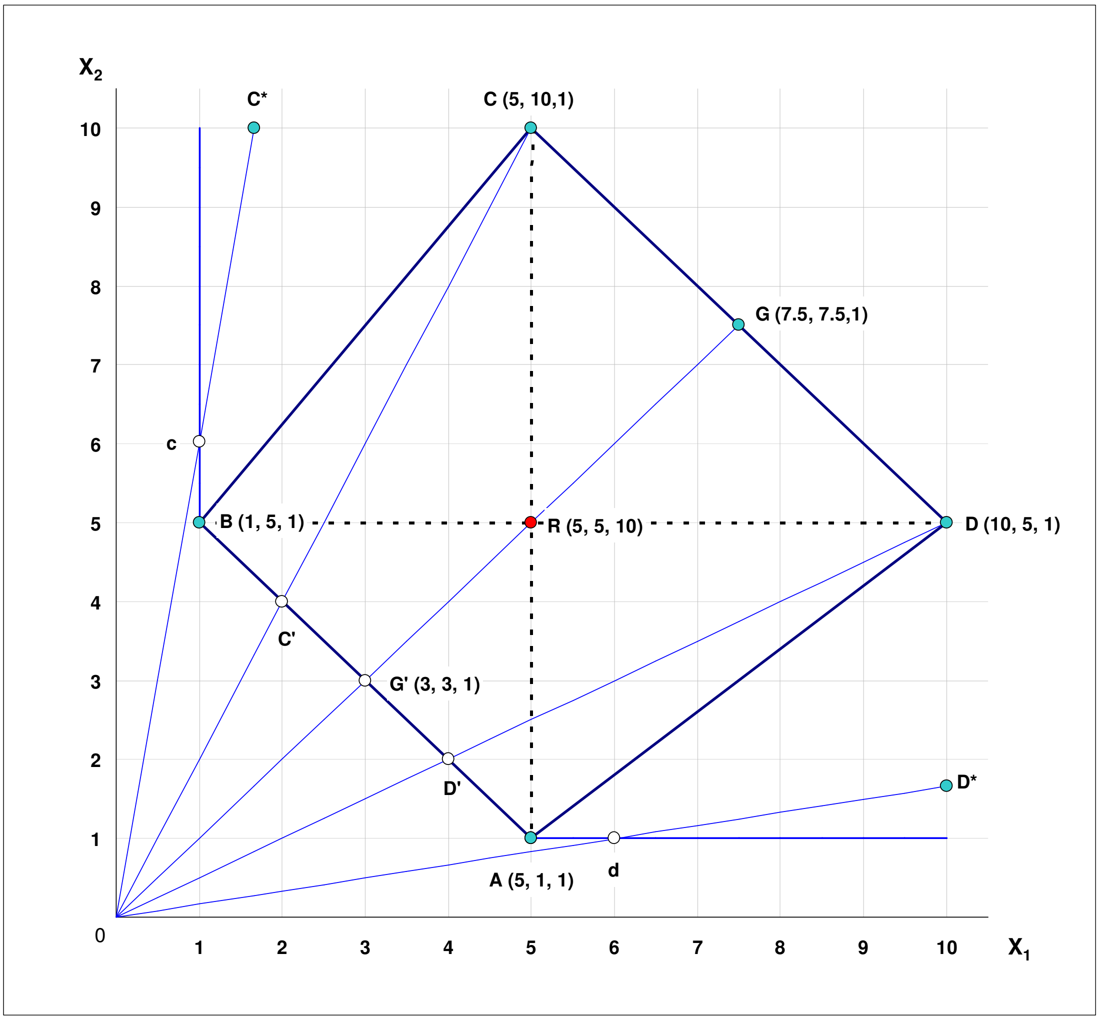

Fig. 3. An illustrative example

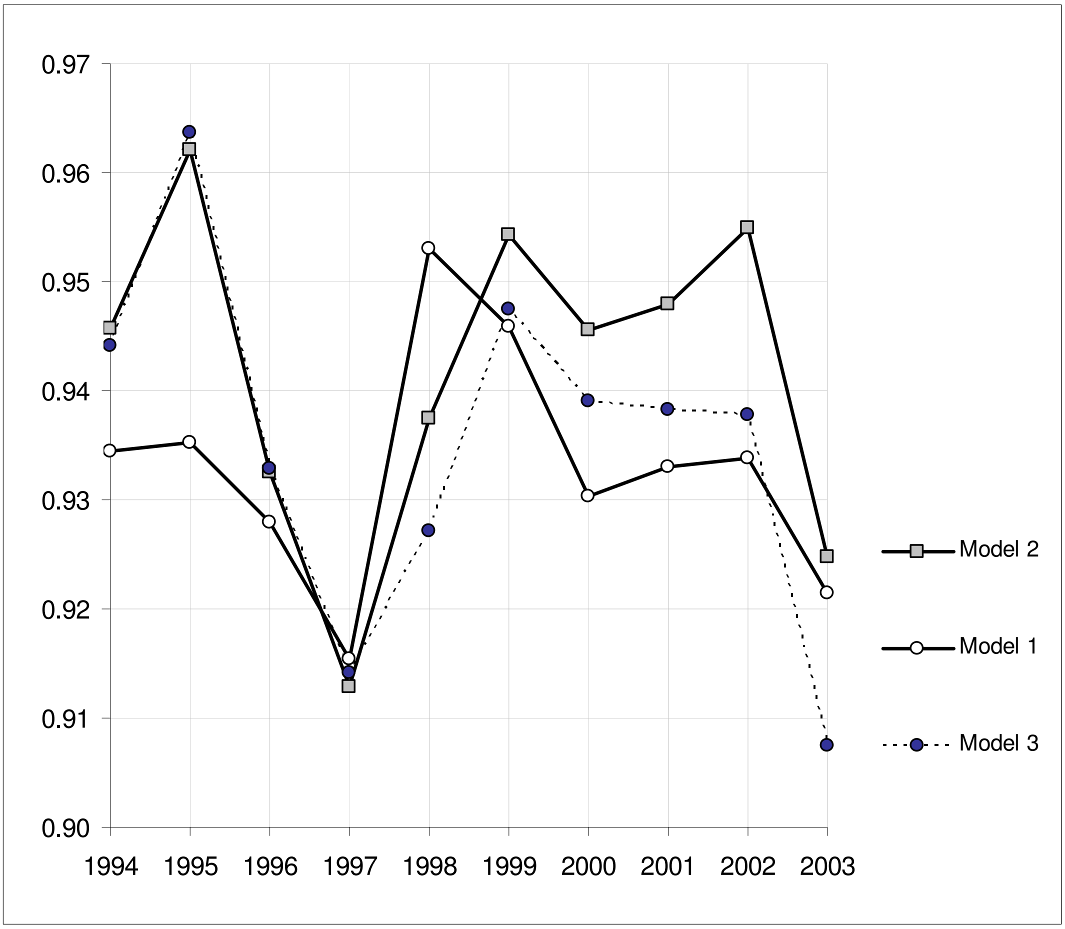

Fig. 4. Unweighted mean TE scores from alternative models (n = 45)

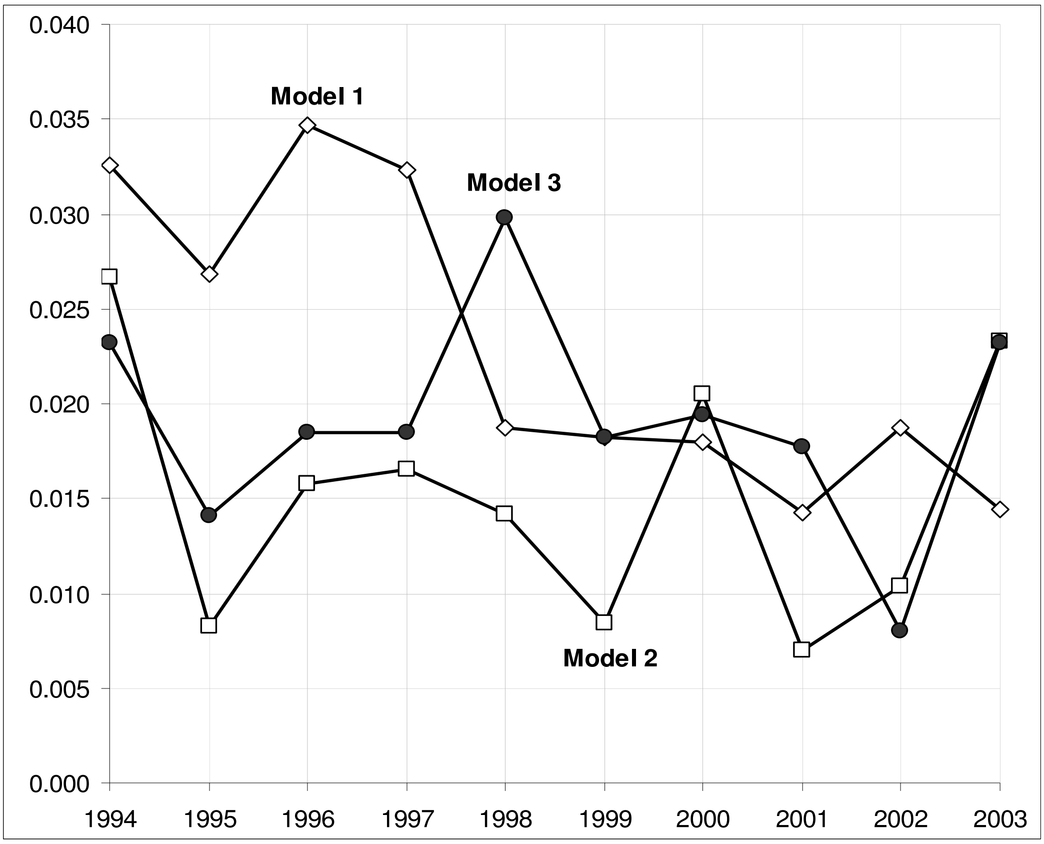

Fig. 5. Cooper's approach: mean congestion scores from alternative models (unweighted, n = 45)

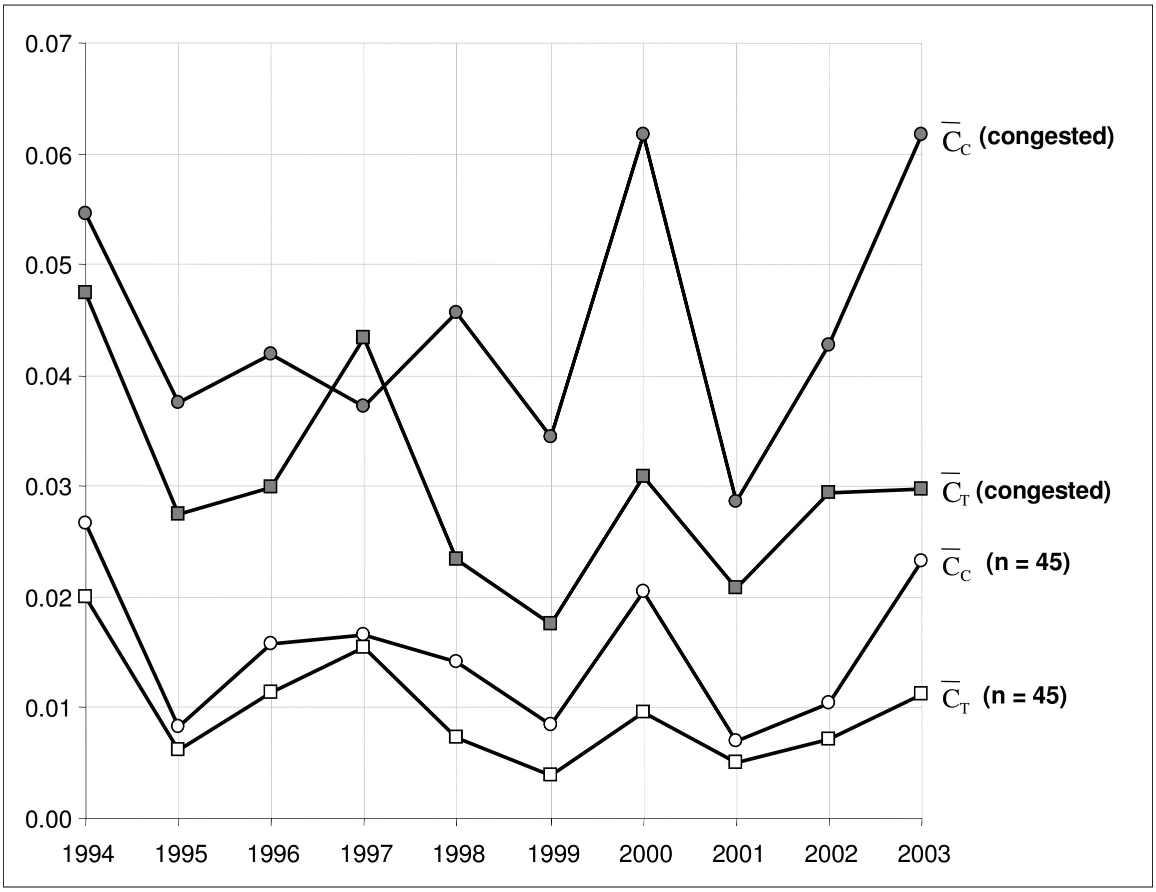

Fig 6. Cooper's approach versus Tone's approach: mean congestion scores from alternative models (unweighted, n = 45)

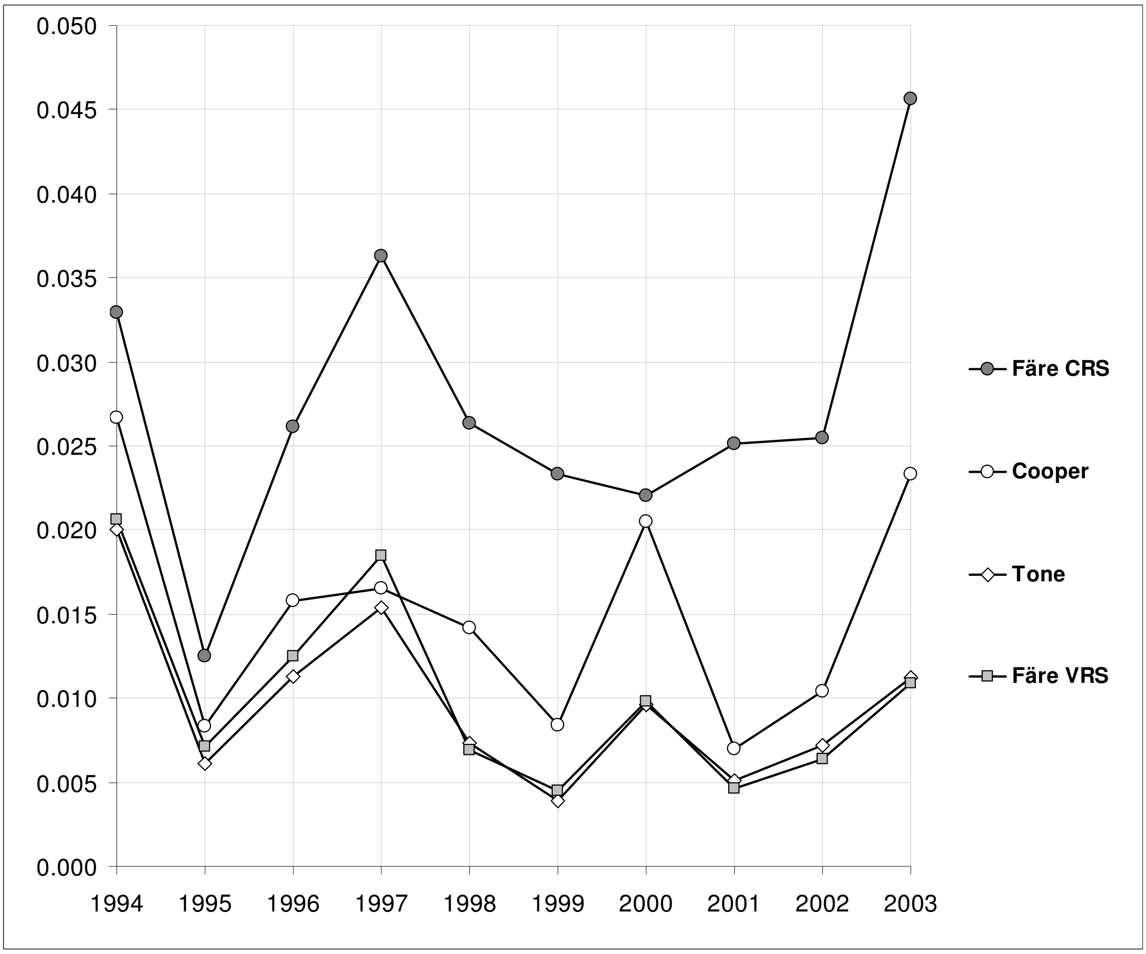

Fig. 7. Mean congestion scores from alternative approaches (unweighted, n = 45)

Table 1. Annual mean TE scores for alternative models (n = 45)

	TE (UAM)	TE (WAM)	Difference	Min	SD	No. on frontier
Model 1						
1994/5	0.934	0.923	–0.011	0.764	0.075	19
1995/6	0.935	0.923	–0.012	0.783	0.066	14
1996/7	0.928	0.917	–0.011	0.753	0.084	18
1997/8	0.915	0.914	–0.002	0.758	0.078	12
1998/9	0.953	0.957	0.004	0.722	0.061	18
1999/0	0.946	0.948	0.002	0.747	0.065	17
2000/1	0.930	0.927	–0.003	0.731	0.082	18
2001/2	0.933	0.930	–0.003	0.720	0.079	21
2002/3	0.934	0.931	–0.003	0.736	0.076	17
2003/4	0.921	0.931	0.010	0.682	0.092	16
Model 2						
1994/5	0.946	0.944	–0.001	0.785	0.061	18
1995/6	0.962	0.958	–0.004	0.846	0.047	22
1996/7	0.933	0.933	0.001	0.728	0.073	16
1997/8	0.913	0.923	0.010	0.734	0.080	13
1998/9	0.937	0.947	0.009	0.719	0.074	15
1999/0	0.954	0.961	0.007	0.813	0.057	18
2000/1	0.946	0.947	0.001	0.799	0.059	15
2001/2	0.948	0.950	0.002	0.772	0.067	19
2002/3	0.955	0.960	0.005	0.766	0.061	20
2003/4	0.925	0.937	0.012	0.747	0.073	12
Model 3						
1994/5	0.944	0.951	0.007	0.800	0.062	17
1995/6	0.964	0.963	–0.001	0.830	0.047	21
1996/7	0.933	0.941	0.008	0.753	0.067	15
1997/8	0.914	0.928	0.014	0.736	0.079	14
1998/9	0.927	0.939	0.012	0.719	0.070	14
1999/0	0.947	0.954	0.007	0.796	0.061	19
2000/1	0.939	0.941	0.002	0.772	0.068	14
2001/2	0.938	0.937	–0.001	0.713	0.070	17
2002/3	0.938	0.947	0.009	0.667	0.077	16
2003/4	0.907	0.923	0.015	0.695	0.086	11

Table 2. Cooper's congestion scores for alternative models

All 45 universities					Congested universities	
	$\overline{C}_C$ (UAM)	$\overline{C}_C$ (WAM)	Difference	SD	Number	$\overline{C}_C$ (UAM)
Model 1						
1994/5	0.0326	0.0287	–0.0039	0.0445	21	0.0700
1995/6	0.0268	0.0250	–0.0018	0.0412	20	0.0603
1996/7	0.0347	0.0335	–0.0012	0.0507	19	0.0821
1997/8	0.0323	0.0295	–0.0028	0.0488	22	0.0660
1998/9	0.0187	0.0172	–0.0015	0.0354	14	0.0602
1999/0	0.0182	0.0187	0.0005	0.0431	11	0.0743
2000/1	0.0180	0.0173	–0.0007	0.0389	13	0.0623
2001/2	0.0143	0.0142	–0.0001	0.0297	13	0.0497
2002/3	0.0187	0.0175	–0.0012	0.0330	15	0.0562
2003/4	0.0144	0.0127	–0.0017	0.0309	13	0.0499
Model 2						
1994/5	0.0267	0.0235	–0.0032	0.0378	22	0.0546
1995/6	0.0083	0.0088	0.0005	0.0215	10	0.0374
1996/7	0.0158	0.0156	–0.0002	0.0291	17	0.0418
1997/8	0.0165	0.0156	–0.0009	0.0258	20	0.0372
1998/9	0.0142	0.0131	–0.0011	0.0284	14	0.0456
1999/0	0.0084	0.0078	–0.0006	0.0197	11	0.0344
2000/1	0.0205	0.0186	–0.0019	0.0386	15	0.0616
2001/2	0.0070	0.0064	–0.0006	0.0186	11	0.0286
2002/3	0.0104	0.0080	–0.0024	0.0226	11	0.0427
2003/4	0.0233	0.0204	–0.0019	0.0374	17	0.0617
Model 3						
1994/5	0.0232	0.0174	–0.0058	0.0401	18	0.0579
1995/6	0.0141	0.0123	–0.0018	0.0287	17	0.0373
1996/7	0.0185	0.0187	0.0002	0.0343	18	0.0463
1997/8	0.0185	0.0177	–0.0008	0.0288	20	0.0416
1998/9	0.0298	0.0251	–0.0047	0.0371	21	0.0639
1999/0	0.0182	0.0166	–0.0016	0.0321	16	0.0512
2000/1	0.0194	0.0178	–0.0016	0.0340	17	0.0514
2001/2	0.0177	0.0178	0.0001	0.0287	17	0.0468
2002/3	0.0080	0.0065	–0.0015	0.0184	14	0.0257
2003/4	0.0232	0.0204	–0.0028	0.0324	20	0.0522

Table 3. Results from Tone's approach and comparison with Cooper's approach (Model 2, unweighted)

All 45 universities				Congested universities				
	$\overline{C}_C$	$\overline{C}_T$	Difference	Number	$\overline{C}_T$	$\overline{\rho}$	Max	Min
1994/5	0.0267	0.0200	0.0067	19	0.0474	−1.55	−5.98	−0.07
1995/6	0.0083	0.0061	0.0022	10	0.0275	−1.28	−3.42	−0.23
1996/7	0.0158	0.0113	0.0045	17	0.0299	−2.15	−17.40	−0.07
1997/8	0.0165	0.0154	0.0011	16	0.0433	−1.42	−3.08	−0.38
1998/9	0.0142	0.0073	0.0069	14	0.0235	−1.04	−2.35	−0.20
1999/0	0.0084	0.0039	0.0045	10	0.0176	−1.06	−2.32	−0.30
2000/1	0.0205	0.0096	0.0109	14	0.0309	−3.53	−28.34	−0.30
2001/2	0.0070	0.0051	0.0019	11	0.0209	−1.86	−9.76	−0.04
2002/3	0.0104	0.0072	0.0032	11	0.0295	−1.22	−2.25	−0.26
2003/4	0.0233	0.0112	0.0121	17	0.0296	−1.34	−3.54	−0.31
Mean	0.0151	0.0097	0.0054		0.0300	−1.65		

Table 4. Results from Färe's approach and comparison with approaches of Cooper and Tone (Model 2, unweighted, n = 45)

	$\overline{C}_{F,\,VRS}$	$\overline{C}_{F,\,VRS} - \overline{C}_T$	$\overline{C}_{F,\,VRS} - \overline{C}_C$	$\overline{C}_{F,\,CRS}$	$\overline{C}_{F,\,CRS} - \overline{C}_T$	$\overline{C}_{F,\,CRS} - \overline{C}_C$
1994/5	0.0206	0.0006	−0.0061	0.0329	0.0129	0.0062
1995/6	0.0071	0.0010	−0.0012	0.0125	0.0064	0.0042
1996/7	0.0125	0.0012	−0.0033	0.0261	0.0148	0.0103
1997/8	0.0185	0.0031	0.0020	0.0363	0.0209	0.0198
1998/9	0.0069	−0.0004	−0.0073	0.0264	0.0191	0.0122
1999/0	0.0045	0.0006	−0.0039	0.0233	0.0194	0.0149
2000/1	0.0098	0.0002	−0.0107	0.0220	0.0124	0.0015
2001/2	0.0047	−0.0004	−0.0023	0.0252	0.0201	0.0182
2002/3	0.0064	−0.0008	−0.0040	0.0255	0.0183	0.0151
2003/4	0.0109	−0.0003	−0.0124	0.0456	0.0344	0.0223
Mean	0.0102	0.0005	−0.0049	0.0276	0.0179	0.0125

Table 5. Correlations: Model 2, n = 450

	C_T	C_C	$C_{F, CRS}$
C_C	0.647		
$C_{F, CRS}$	0.419	0.328	
$C_{F, VRS}$	0.944	0.651	0.391

Table 6. Percentage contribution of each input to congestion in congested universities

	Other expenditure	Academic staff	Postgrads	Undergrads	Number congested	$\overline{C}_C$ (UAM)
Model 1						
1994/5	12.3	21.8	30.4	35.6	21	0.0700
1995/6	2.8	27.5	29.5	40.2	20	0.0603
1996/7	4.2	28.0	23.9	43.9	19	0.0821
1997/8	7.0	22.1	31.3	39.6	22	0.0660
1998/9	8.6	19.2	19.3	53.0	14	0.0602
1999/0	8.0	9.3	35.6	47.1	11	0.0743
2000/1	2.9	9.1	51.8	36.2	13	0.0623
2001/2	5.6	12.9	31.0	50.5	13	0.0495
2002/3	6.5	19.5	18.6	55.4	15	0.0562
2003/4	2.6	17.5	32.9	47.0	13	0.0499
Mean	6.1	18.7	30.4	44.8		0.0679
Model 2						
1994/5	17.7	27.4	31.8	23.1	22	0.0546
1995/6	15.0	19.7	61.1	4.1	10	0.0374
1996/7	12.5	28.4	36.3	22.8	17	0.0419
1997/8	10.7	23.2	35.9	30.2	20	0.0372
1998/9	11.3	5.9	33.5	49.3	14	0.0456
1999/0	14.5	0.0	41.1	44.4	11	0.0344
2000/1	2.1	19.6	53.1	25.2	15	0.0616
2001/2	14.1	0.0	65.5	20.5	11	0.0286
2002/3	19.5	12.6	49.1	18.8	11	0.0427
2003/4	7.6	15.2	40.4	36.9	17	0.0617
Mean	12.5	15.2	44.8	27.5		0.0447
Model 3						
1994/5	25.6	32.2	25.5	16.8	18	0.0579
1995/6	25.9	19.8	43.9	10.4	17	0.0373
1996/7	13.5	20.4	35.4	30.7	18	0.0463
1997/8	30.3	9.9	27.2	32.6	20	0.0417
1998/9	26.1	5.2	32.9	35.8	21	0.0639
1999/0	24.0	2.9	54.5	18.5	16	0.0512
2000/1	14.4	1.6	51.3	32.7	17	0.0514
2001/2	26.1	4.0	34.4	35.6	17	0.0468
2002/3	12.9	9.2	50.0	27.9	14	0.0257
2003/4	5.4	20.0	36.4	38.1	20	0.0522
Mean	20.4	12.5	39.2	27.9		0.0500