FEATURE

Following the Atkinson Review: the quality of public sector output

Martin Weale
National Institute of Economic and Social Research

SUMMARY

The UK Centre for the Measurement of Government Activity has made substantial progress in the measurement of public service output and inputs, publishing a series of productivity articles, but inevitably there is room for more work, particularly in the measurement of the quality of both outputs and inputs.

This article addresses two of the key issues raised by the Atkinson Review – quality adjustment and the use of value weights. The benefit of addressing quality issues in the context of conventional index number formulae is shown. This leads to smaller quality adjustments than some past work on public sector output has suggested. It is demonstrated that, without the use of value weights, it is not always possible to make quality adjustments. Nevertheless, where value weights cannot be based on market information, they may be difficult to identify and care will be needed in identifying changes to relative values.

Measurement of public sector activity in constant prices poses problems with no straightforward solutions. For many years it seemed to be the Cinderella of national income accounting. In the 1990s, considerable effort was devoted to the production of price and volume indices for the information technology sector, because of a feeling that simple price and volume indices greatly understated the growth of the industry. The public sector is considerably larger than the information technology sector. However, after some experimentation with activity measures, from the 1960s until the 1990s outputs were typically measured by means of rather crude indicators of inputs, such as numbers of people employed.

The 1993 System of National Accounts, followed by the 1995 European System of Accounts, proposed a move away from input indicators to activity indicators. Instead of counting the number of teachers in schools, one should count the number of children being taught. Health output might be measured by the number of patients treated and not the number of people employed by the health service. The UK, often in the lead in implementing new national accounting standards, started to move towards output-based measures in 1998. The effect was to depress estimates of economic growth to some extent.

There was a large increase in public spending between 1999 and 2005 and, at much the same time, the Government adopted a system of targets for public sector services in order to monitor and, it was hoped, improve performance. It was therefore of particular concern that the National Accounts showed labour productivity in the public sector declining. In 2003 the National Statistician asked Sir Tony Atkinson to review the problems of measuring the output, and thus the productivity, of the public sector. The Review led to the setting up of the UK Centre for the Measurement of Government Activity (UKCeMGA) at the Office for National Statistics, as a way of implementing its proposals.

The Review produced nine principles. Key to these were the first two – that, as far as possible, public sector outputs should be treated in the same way as private sector outputs and that adjustments should be made for changes in quality. The Review also proposed that indices of individual components of output should be weighted together using value weights rather than cost weights. After summarising some of the work done since UKCeMGA was set up, this article focuses on these two intimately related questions.

Progress so far

UKCeMGA has looked so far at four areas: education, health, adult social care and social security administration. Its papers on the first two areas include estimates of the productivity performance of the sectors after making quality adjustments, while the papers on the final two areas discuss a number of possible indicators but do not yet provide any quantitative assessment of their implications. A strategy paper (ONS,

2007) published on 3 July 2007 consolidates this work and sets out plans for the future. It also provides details of specific decisions taken on measurement methods for health and education.

Measurement of productivity involves measurement of inputs and outputs and the contribution made by labour and capital. Construction of volume measures of value added requires the first two of these, although the UK has traditionally short-cut the issue by assuming that volume movements in gross output match volume movements in value added.

UKCeMGA has made substantial progress in the measurement of all of these questions, focusing on activity measures. This has largely involved increased distinction between the different types of activities carried out by components of the public sector. Thus, before there was substantial concern about volume measures for public sector activities, hospital output was measured by classifying each procedure into one of 16 activities. It is easy to believe that such a crude classification could result in substantial biases, and that a move to an index of some 1,600 activities is an improvement. On the other hand, care is needed to ensure that all the output of any sector is enumerated and this has led to a preference for calculating output indices for the market sector by deflating value measures rather than by collecting output volumes. The risk of activity-based measures neglecting new forms of activity may increase as the degree of disaggregation is increased.

Considerable effort has also been devoted to the measurement of factor inputs. Here, as in any productivity calculation, it is important to measure labour input after allowing for differential use of different types of labour rather than simply using a head count or a measure of total hours worked. If this is not done, changes in total factor productivity will be confused with changes in labour input. Similarly, attention needs to be given to the measurement of capital services, rather than reliance put on indices of the capital stock. In both of these areas UKCeMGA has achieved a great deal.

Inevitably, however, there is room for more work in the measurement of the quality of both outputs and, particularly in the case of the health service, inputs. The question of outputs is discussed below. On the issue of inputs, it is adequate to note that there has been substantial technical progress in the pharmaceutical industry in terms of capabilities of drugs available. There has, so far, been no attempt to produce quality-adjusted measures of the output of pharmaceutical industry and thus of the inputs bought in by the health service.

Table 1
Productivity estimates for components in the public sector, 2004

Indices (1999=100)

	Consistent with current National Accounts	After quality adjustment
Education	90	100
Health	93	98
Social security administration	82	
Adult social care	92	

Table 1 shows the estimates of changes in public sector productivity produced by UKCeMGA after taking account of the quality effects which it has been able to identify. The results are only approximate, because the presentation of the estimates in the various papers from which they are drawn, and for which references are given in the table, is graphical. Thus, the numbers for 1999 and 2004 have to be read off the graphs. Nevertheless, the broad impression they give is adequate.

It has to be said that the results do not provide a flattering picture of public sector productivity performance. In fields of both education and health, the quality adjustments have made a substantial difference; there are, nevertheless, reasons for questioning whether the adjustments made have been appropriate and this issue is now explored.

Quality adjustment

The basic principle behind quality adjustment of the output of the public sector is the same as that elsewhere in the National Accounts – outputs of different quality should be treated as distinct outputs in the construction of quantity indices of output. Suppose that there are two qualities of output in period t, q_{1t} and q_{2t} with base-period unit values p_1 and p_2. Then the quality-adjusted Laspeyres output index comparing period t with period 1 is

$$I_t^q = \frac{p_1 q_{1t} + p_2 q_{2t}}{p_1 q_{11} + p_2 q_{21}} \times 100$$

This compares with the unadjusted index calculated if the output is assumed to be homogeneous

$$I_t^h = \frac{q_{1t} + q_{2t}}{q_{11} + q_{21}} \times 100$$

The same principle can of course be applied to the calculation of other types of index, such as the chain-linked index which is actually used in the National Accounts nowadays.

A question to which the formula immediately gives rise is where the unit values come from to make the calculation possible. If output were marketed, they would of course be the market prices (or the market prices net of sales taxes). In the absence of a market, the most obvious choice is that the social values of the different outputs should be used. In some cases this may be very straightforward, and an application to education illustrates this very clearly.

Education

Suppose that the question is how to produce an index of education output which reflects changes in the quality of teaching and that this can be measured by exam results. Two quality categories are identified, children with 5+ GCSEs at grades A to C and those who have not reached this level. Suppose that the difference in the unit values of the two qualities of education is reflected in differences in earning power. Suppose also that, on average, a child with 5+ GCSEs earns 20 per cent more than one who does not cross this threshold (this is consistent with the figures produced by McIntosh (2006), although he identifies separately the value of lower levels of GCSE attainment). Then p_1 =1 and p_2 =1.2. If the proportion of children with 5+ GCSEs rises from 60 per cent to 63 per cent and the number of children is unchanged, it can be seen immediately that

$$I_t^q = \frac{1 \times 37 + 1.2 \times 63}{1 \times 40 + 1.2 \times 60} \times 100 = 100.5$$

while the homogeneous index shows no change. The outcome can also be compared with that proposed by the Department for Education and Skills (DfES, 2005), which recommends an index

$$I_t^d = \frac{63}{60} \times 100 = 105$$

Comparison of that with the formula for I_t^q shows that the latter would take a value of 105 only if the value put on children with

no qualifications were zero or, which is an easier proposition to defend, if it were believed that their education had no effect on their earning power (something not supported by McIntosh's results since he shows earnings benefits from some success at GCSE for children who do not reach the 5 A to C grade threshold). While cases can be made for other valuation systems, the use of figures based on what is known about earning power is not likely to be controversial.

These calculations are not the whole of the matter since the exam scores of children currently taking GCSEs are presumably a function of the quality of their teaching throughout their time at school. Thus, there are serious issues to be resolved about the best way of allocating the quality effects over time. ONS (2007) sets out a programme of work to address this issue. But the general principle that quality adjustment can be seen as an index number issue is clear enough.

One final point should be made since there is some discussion among national accountants about whether value weights are more appropriate than cost weights (see below). In this example at least, cost weights would seem to be highly inappropriate. It is perfectly possible that teaching children who reach 5+ GCSEs is no more expensive than teaching children who do not cross the threshold. But this hardly implies that the extra education implied by the better exam result is of no use. One might also note that the use of value weights defined in this way is entirely consistent with the way in which labour force quality is measured in productivity calculations. On the other hand, one can reasonably be concerned if post-compulsory education is not valued unless the latter leads to enhanced earning power, otherwise it would neglect the consumption value of such courses and the knowledge derived from them to those who undertake them.

Ideally, instead of categorising children, the exam score of each would be identified and a measure of quality built from this. ONS (2007) proposes this approach in preference to the threshold measure described above. To apply the above principles, a unit value would need to be associated with each exam score. In practice, of course, it is unlikely that unit values will be able to be identified in this way.

Suppose, however, that the unit values are given as linear functions of the exam marks and the value p_s associated with a score of s is given as

$$p_s = a + b_s$$

Then, if $q_{s,t}$ is the number of children with score s in year t, the quality-adjusted index is

$$I_t^q = \frac{\sum_s (a+bs)q_{s,t}}{\sum_s (a+bs)q_{s,0}} \times 100 = \frac{(a+b/\bar{s}_t)\sum_s q_{s,t}}{(a+b/\bar{s}_0)\sum_s q_{s,0}} \times 100$$

where $\bar{s}_t$ is the average score in year t. The important point to note about this is that if $a>0$, a 1 percentage point improvement in exam score is associated with a less than 1 percentage point improvement in earning power. The use, proposed by ONS (2007), of ratio of average marks in different years as a means of quality adjustment will overstate the improvement which would be shown by the appropriate index number. (It is possible to imagine $a=0$ but harder to accept $a< 0$. The latter would imply that school subtracted value from children who performed badly in exams.) Unless there is a firm statistical basis for the function $p_s = a+bs$, it may well be better to rely on a rather small number of categories for which pay differentials and thus unit values can be derived. It should also be noted that if the value function is not linear in the exam score, then use of the mean is doubtful.

Health

The study of hospital output by Castellani *et al* (2007) provides another illustration of the problems arising in making quality adjustments. A reduction in the mortality rate associated with hospital treatment is plainly an improvement in quality. At present, for most in-patient treatments, it is possible to distinguish only two categories, patients who survive and those who do not. Values can be obtained for the two types of outputs by using, for the surviving patients, a measure of the increase in quality-adjusted life years and, for the dying patients, the quality-adjusted life years lost as a result of their treatment. Since there is very little information available on the benefits of different treatments, it is possible only to make rather arbitrary assumptions about the gains relative to the losses.

There are a number of difficult issues. If a treatment raises the welfare of a patient by a uniform amount for each remaining year of life, is the treatment of a young patient 'more output' than that of an old patient? It may seem sensible to treat the deaths of young patients as bigger losses than the deaths of old patients, but to assume that, for surviving patients, the amount of treatment is not dependent on their ages. On the other hand, if hospital treatment is seen as saving people who would have otherwise died, then no value is actually subtracted by patients' deaths. Ideally, as proposed by ONS (2007), one would make a distinction between those patients who die as a result of their treatment (avoidable death) and those who die simply because their treatment does not work (unavoidable death). Of course, where the patient is provided with terminal care because nothing else can be done, that is in itself valuable and should not be treated in the same way as other unavoidable death.

The Department of Health (DH) has been keen to ensure that an index of output pays due regard to the quality of the patient experience with respect to the hotel services offered by hospitals. Here, problems arise similar to those involved in the measurement of school quality. Patient experience is measured by means of sample surveys of patients who are asked to report on a range of issues such as food quality, hospital cleanliness and staff politeness. The question then is how the unit values for various treatments should be adjusted upwards or downwards in the light of the patient scores.

Suppose that a particular treatment is identified (such as a hip replacement) with a unit value of p_j in terms of the medical benefits conferred by the treatment. The patient has also given a score of s for the non-medical aspects of the treatment. What then is the total value of the treatment to the patient and how would it change if the quality of the non-medical aspects changed? Suppose that one could identify a value to be put on the score, $v_j=a_j s$ (although there is no need for the relationship to be linear). The value on the score may well depend on the procedure in question, for example because the value put on a short waiting time is likely to depend on what the patient is waiting for.

This suggests a score-adjusted value of the treatment as

$$p_{j,s} = p_j + a_j s$$

where p_j is the unit value for treatment j for a surviving patient. It is plain that to apply this formula one has to decide, somehow or other, on the value for a_j. The measure adopted by DH avoided this problem by assuming that

$$p_{j,s} = sp_j$$

but it is very difficult to see a justification for this approach beyond the point that it avoids the need to take a view on what is in fact the key issue, the importance of non-medical quality relative to medical treatment, in treatment packages. ONS (2007) suggests that this issue will now be addressed.

These examples show the importance, when making quality adjustments, of starting with the basic principles lying behind index numbers. This leads naturally to the treatment of different qualities of output as output of different products with different unit values attached. It also leads to a consideration of the effects of quality on unit values. It is unlikely from work to date, either by the various government departments or, indeed, in the treatment of patient experience in the York/National Institute of Economic and Social Research study of hospital output, that the quality adjustments were approached from this perspective.

The issue of quality can blur into that of new products. For example, if a form of medical treatment is improved in terms of what it does for patients, then the treatment can be regarded as a new product with a unit value higher than that of the treatment it replaced. Provided that the unit value can be identified, then the calculation of the output index is quite straightforward.

Valuation issues

The Atkinson Review proposed that, where possible, value weights rather than cost weights should be used in the construction of the output indices described above. Thus, if the value of some particular product or activity to the private sector is believed to be higher than its cost of production, the importance of changes in this activity in an overall output measure will be enhanced compared with what would be produced by a cost-based output index.

The logic behind this is clear. In the private sector, an output index is calculated using the prices at which the goods produced are then sold, and not on the basis of their cost of production. The two are conventionally equated, because profit is calculated as the residual, although it is regarded as the cost of capital. Thus the calculation is clearly driven by sales values rather than any identifiable costs of production.

The National Accounts, at present, have the property that value of output equals cost of production, with profit being treated as a cost of production. This identity is the core of the accounting system. For the private sector, since profit is calculated as a residual by deducting other production costs from output value, the identity will always be met. But in the public sector, if output values are allowed to differ from costs of production, the key identity will be broken. Alternatively, some new income category will be needed to maintain the income/expenditure identity which is core to current price National Accounts. Since the residual term in the existing National Accounts is called operating surplus, a possible name for the new type of income might be social surplus.

In some cases the derivation of the values of outputs may be reasonably straightforward. Thus, in the education example discussed above, there is a market for qualified labour and that market provides a measure of the value offered by education. In other cases there is no market, and appropriate means are needed to assess the values that people put on public services. Where this is possible, care is needed to deal with the point that different people value services differently. In the private sector, those consumers who value a service at or above its sale price buy the service, while those who value it less do not. The National Accounts do not measure the consumer surplus associated with the people who would be prepared to pay higher prices if they had to.

Many public sector outputs are goods or services which are provided because of need. Healthy people do not expect hip replacements and, of those who are treated, the expected benefit to some is probably greater than to others. Thus, the principle of marginal valuation needs to be amended to reflect the fact that, for many people, the good or service may be of no use at all. To apply the private sector principle, the service provided by the public sector should be valued on the basis of the value put on it by the marginal consumer, that is, the consumer who derives the least benefit from the treatment. Thus, if health treatment is valued more by a young person than an old one, because the former has a longer expected life, the social value would be given by the value of the treatment to the old person. In the same way, police protection might be valued more by high earners than low earners because the former may put higher values on their own lives. But the value used by the low earners would be that adopted in the National Accounts.

Illustration of the way in which this principle works clarifies a number of points which have confused at least this author. Suppose that, for reasons nothing to do with the health service, people live longer. The value they put on treatments such as cataract operations is likely to increase because they will be in a position to enjoy the benefits for longer. If the cost is unchanged and allocation is efficient, the number of people treated will rise to the point where the benefit enjoyed by the marginal patient is what it was before.

Has the volume index of cataract operations increased in line with the number of patients treated? Or has it increased more because the non-marginal patients are enjoying more quality-adjusted life years – a widely accepted measure of health outputs? The value of the treatment to the marginal patient does not change, even though more patients are treated. Marginal valuation therefore implies that changes in the number of operations (the activity measure) are therefore the appropriate guide to the change in output. Of course there is no guarantee that allocation of resources to cataract operations is efficient. But statisticians would probably be unhappy with a departure from the principles set out above on the grounds that too few or too many patients are treated. Should the value of the treatment to the marginal patient change because more or fewer patients are treated, that seems *prima facie* to be a price change rather than a quantity change. In practice, local variations in health services mean that a patient who would be treated in one area may be refused treatment in another area. Addressing this needs further thought but, in the short term, national accountants would probably feel most comfortable using the lowest of a possible range of marginal values, when these differ for geographic reasons.

This can be contrasted where the number of quality-adjusted life years associated with a medical procedure changes because of changes to the procedure or some other technical advance. In that case, the normal process would be to treat the improvement to the procedure as an improvement in quality which should be reflected in a volume index. Here, the proportionate increase in volume is probably best derived with reference to the patient who was marginal under the old procedure. The ratio of quality-adjusted life years for this patient under the new procedure relative to that under the old procedure gives the proportionate increase in volume. Thus, if the old procedure had not been applied to patients over the age of, say, 75 because of its cost, while the new procedure was worth providing to older patients, the volume index is derived from consideration of the 75 year old patient.

These principles no doubt need further elaboration. But the examples above demonstrate the importance of thinking about the impact on the marginal beneficiary of public spending. They also show that, while it is important to think about outcomes as well as activities, there are circumstances in which changes in activity measures are more appropriate

than changes in outcome measures as indicators of output movement. Perhaps they also point to a principle which would probably be widely accepted that, if nothing happens in the production process of a public service, then the output index should remain unchanged even if, for some completely exogenous reason, the benefit derived from the service to the marginal consumer increases.

Conclusions

UKCeMGA has achieved a great deal in enhancing measures of input and output for individually consumed public services. Progress with collectively consumed services is likely to be harder. Even with individual services, as the discussion above suggests, there remains a considerable amount of conceptual clarification to be done there is a substantial amount of work to do in the way in which quality changes are treated and the author looks forward to seeing more progress in this area.

CONTACT

elmr@ons.gsi.gov.uk

REFERENCES

Castelli A, Dawson D, Gravelle H, Jacobs R, Kind P, Loveridge P, Martin S, O'Mahony M, Street A, Stevens P, Stokes L and Weale M R (2007) 'A New Approach to Measuring Health System Output and Productivity', *National Institute Economic Review* No 200.

Department for Education and Skills (2005) 'Measuring Government Education Output in the National Accounts' at www.dfes.gov.uk/research/programmeofresearch/index.cfm?type=5

McIntosh S (2006) 'Further analysis of the returns to academic and vocational qualifications', *Oxford Bulletin of Economics and Statistics* Vol 68 pp 225–51.

Office for National Statistics (2007) 'Measuring Quality as a Part of Public Service Output' at www.statistics.gov.uk/cci/article.asp?id=1831