

**MEASUREMENT OF UK GOVERNMENT OUTPUT AND PRODUCTIVITY FOR THE
NATIONAL ACCOUNTS**

A. B. Atkinson
Nuffield College, Oxford

(read before the Society, 19th May 2005)

Abstract: The Atkinson Review was commissioned by the UK National Statistician in December 2003 to advance methodologies for the measurement of government output, productivity and associated price indices in the context of National Accounts. The report sets out the principles that should underpin future work in the measurement of government output and productivity. It makes specific proposals for improving output measures in Health, Education, Public Order and Social Protection. It also recommends broad approaches for developing the measurement of the output of public services that will improve the accuracy, coverage and interpretation of figures, with publication of methods and figures to ensure transparency and debate on ways of achieving further improvements. Sir Tony Atkinson, Warden of Nuffield College, Oxford, is one of the worlds leading economists and an expert in public economics.

Keywords: Government Output, Productivity, Public Services
JEL Classification: H42, H5

INTRODUCTION

At the end of January 2005, I published the Final Report (Atkinson, 2005) of an independent review of the measurement of government output and productivity in the United Kingdom (UK) national accounts. The main conclusions of the yearlong review may be summarised as follows:

- “Public accountability requires measures of what is achieved by spending on public services.”
- “We cannot simply assume that outputs equal inputs in such a major part of the economy, which is a fifth of UK Gross Domestic Product.”
- “The UK Office for National Statistics has taken the right route in introducing direct measures of government output, but the statistics on government activity need to be developed further, in a principled framework.”

The report made 54 recommendations as to how to improve the measurement of public services output and productivity. Work had already been undertaken by the Review Team (led by Joe Grice and Aileen Simkins) to develop new government statistics on health, education, social protection, and criminal justice. The then National Statistician, Len Cook, responded very positively to the Report by announcing the creation of a new UK Centre for the Measurement of Government Activity within the Office for National Statistics (ONS) to carry forward the work. This new Centre (CeMGA) was established in July 2005 under the direction of Joe Grice.

In Section 1 of this paper, I describe the background to the Review, with a particular emphasis on the extent to which the issues are of general, rather than purely UK, interest. In Section 2, I summarise the main problems that the Review identified with the present statistical methods used in the UK. Consideration of these problems led to the Report making recommendations for a wide range of changes. Here I concentrate, not on the individual proposals, but on the underlying principles, which are the subject of Section 3.

1. BACKGROUND TO THE REVIEW

What was the origin of the Review? Obviously public spending and the public sector are very much in the public eye. But the Review had a particular focus. I was asked by the UK National Statistician to advise him on the measurement of government output and productivity *in the national accounts*. This is related to, but different from, measures of performance for the management of the public services. The national accounts are concerned with measuring the overall output of goods and services in the economy, and these are among the most important economic statistics. The total Gross Domestic Product (GDP) is the figure to which reference is made when people talk about “the growth of the UK economy”, and draw comparisons with other countries.

For many years, the contribution of the public sector to national income was measured by simply assuming that output = input. It was taken for granted that you get out what you put in. This assumption is clearly unsatisfactory, since it tells us nothing about how the productivity of the public sector is changing over time. Some countries, such as the US, still measure the output of the government by the inputs, but this is no longer regarded as good statistical practice. More than 10 years ago, in 1993, the UN System of National Accounts (SNA) was revised to recommend that countries drop the assumption that output = input and seek to measure directly the output of public services (United Nations, 1993). The SNA recommendations are not mandatory, and, as just noted, the US is not following them, but in the case of the European Union (EU), they were legally adopted. Under a decision of December 2002, EU Member States are required to introduce direct measures of public sector output in time for the 2006 national accounts (with the exception of Denmark, which secured a derogation until 2012).

Why is the EU concerned? In part, the decision was reached because we rely on national accounts as a means for judging the macro-economic situation. The relationship between input and output in the public sector is a guide to the quality of the public finances of different Member States, and they need to be comparable. Overall, the national accounts are widely used as a measure of economic performance. This is well illustrated with the current obsession with the fact that Europe is growing less slowly than the US. Such a comparison can only be made on the basis of comparable statistics. Given that public output is some fifth of total GDP, its treatment can make a noticeable difference. Indeed in the Report we pointed out that between 1995 and 2003 the US economy grew $\frac{1}{2}$ percent per year faster than that of the UK. The US economy grew at $3\frac{1}{4}\%$ per year, whereas the UK economy grew $\frac{1}{2}$ percent per year less, at $2\frac{3}{4}$ percent. But half this difference is due simply to the difference in how the two countries measure government output. If we had used the same method as the US to measure public output, then the UK growth rate would have been 3 percent per annum. The Eurostat recommendation is intended to eliminate such differences in practice within the EU. (It has no control over the failure of the US to follow international best practice.)

The ONS in the UK was one of the first statistical agencies to respond to the SNA 1993

requirements for measuring government output. The ONS began introducing direct output measures seven years ago. The areas covered are listed in Table 1. On the right hand side are shown the main components of the output indicator. For example, the output of the Department for Work and Pensions in administering social security benefits is measured by the number of benefit claims processed. The output of social services is measured by the number of children and older persons in care. In the second column is listed the total spending in 2000, and this shows that direct measures cover more than 60% of the total general government final consumption.

It is important to make clear the implications for the macro-aggregates. We are concerned here with the measurement of Gross Domestic product at constant prices. This volume measure provides an indicator of the growth of the economy. It is *changes over time* which are captured by the direct output measures; we are not attaching a monetary value to the *level* of government output. The introduction of the direct measures did not change the base statistic. It provided a measure of the productivity change. What this meant in reality over the past ten years is that productivity was recorded first as rising and then as falling. Initially, the inputs were rising slowly, and output was growing faster. Then the government expanded inputs much more rapidly, but the output indicators continued at much the same rate as in the past, so that productivity began to fall.

2. ASSESSMENT OF MEASUREMENT METHODS

The ONS has been a leader in the field, but the methods first applied in 1998 are best described as “first generation” methods, in need of further elaboration. The Review not only recommended major improvements, but also went a significant way, in collaboration with ONS staff, towards developing new approaches.

The first major issue is that of the limited number of the output indicators (identified in the right hand column of Table 1). They were too highly aggregated; not allowing the output measure to capture fully changes in the mix of outputs. In the case of health, for example, the UK Health output measure used before June 2004 was based on only 16 indicator series, a single one of which accounted for about half total covered expenditure. The new measure introduced in the National Accounts published in June 2004 used 1,200 Healthcare Resource Groups and 400 other activity groupings. The current version now covers some 1,900 activities.

The second issue was that coverage was incomplete, as illustrated by the estimates of the output of personal social services. In June 2004, 70% of the weight in the output indicator was attached to the services supplied to older people, whereas they accounted for only 40% of total spending. For children the balance was about right, but that reflected the failure to cover spending on children supported in families or living independently. If they had been included, then children too would have been over-weighted, since services for adults under 65 were not covered. Not only do they account for a sizeable fraction of spending, over 30%, but this means that any shifts in policy priorities will cause the output indicator to be misleading.

Thirdly, coverage was also incomplete in a geographical sense (and remains so). It does seem extraordinary that a volume entitled United Kingdom National Accounts should base its indicators to a substantial extent on Great Britain (excluding Northern Ireland), or England and Wales, or just England. The quality adjustment for education noted in Table 1 was based on examination results in only England. The number of children and adults in care, and the number of hours of home help services, used to construct the personal social services output, were for England only. The experimental measure for the police was based on figures for England and Wales grossed up to cover Scotland and Northern Ireland, and there is no reason to suppose that this is appropriate.

A fourth key message of the Review concerns the importance of improving measures of government inputs as well as outputs. In order to measure the real resources going into the public sector, we need to know how much is being spent and how much any increase is due to inflation rather than a real increase. At an early stage of the Review, we decided to carry out a thorough investigation of the data supply process, which seemed to have been taken largely for granted. Productivity calculations can however go wrong as much through mis-measuring inputs as mis-measuring outputs. An important example is provided by the deflators used to arrive at real inputs. Pay is an important element in spending, but changes in pay rates are treated in a variety of ways, some more satisfactory than others. For example, in some cases, pay is simply deflated by the average earnings index. This has been true of the prison service, the probation service, and the Crown Prosecution Service. There is no reason to suppose that pay has risen at the same rate for all groups, and account has to be taken of pension contributions and National Insurance contributions.

The input side also illustrates a fifth concern, which is with the lack of timeliness. If one tracks the police procurement data from source to the national accounts, one finds that the data come from two streams –the Home Office and the Office of the Deputy Prime Minister – and are subject to a whole series of adjustments and amendments. This means that the process is drawn out. The forms are distributed to police authorities in mid-May, with a deadline of 31 July, but the last forms do not arrive until December, so that while emerging findings are sent to ONS by late November, the final figures are not sent until March. Not only is there little sense of urgency; there is also little sense of purpose. Most importantly the suppliers of the data have little or no idea as to how they are ultimately used. There is no clear line of sight from the initial data supplier to the end use in ONS.

3. RECOMMENDED: A PRINCIPLE APPROACH

The Report did not limit itself to making detailed suggestions for each of the main spending areas. More ambitiously, it set out a principled approach to the measurement of government output and productivity. The nine principles enunciated in the Report, covering the measurement of outputs, inputs, and productivity, underlie the detailed recommendations of the Report, and the checklists that we have given for the criteria to be applied when revising output measures or choosing price deflators.

Although there may not be universal agreement on the principles themselves, I believe that such a principled approach provides an invaluable protection for official statistics in this sensitive area. Changes in the measures of government output should only be introduced where they are in accord with an agreed set of principles. Concerns have been expressed about the introduction of direct output measures on the grounds that such measures would be too exposed to political manipulation. This is a genuine concern, but I see the principles as a means of mitigating this risk. The underlying principles are summarised in Table 2. They vary from straightforward and uncontroversial, such as Principle E on geographical coverage, to the much more complex and potentially controversial. Here, I concentrate on four of the latter. On the output side, there has been much discussion of Principle B, that output should be measured in a way that takes account of changes in quality. In one sense principle B follows from Principle A that the measurement of government output, should, as far as is possible, follow methodology parallel to that appropriate for the private sector. In essence, what we are trying to measure is the same for both sectors. This in turn implies that, just as in the private sector, the measurement of quality is central to our concerns.

I am firmly of the view therefore that measures of government output growth should take account

of quality change. Quality has many dimensions, and some will prove elusive, but there are several possible ways forward. As we describe in the Report, quality change may be treated by distinguishing different products. This is one reason why it was important to extend the degree of differentiation of activities in the health service. Quality change may also be seen as “repackaging”, discussed further below, where the service remains the same but output rises: for example where a higher proportion of letters reach the correct destination on time. Of course, we recognised that making adjustment for quality change is not easy, and may take some time. We were aware that we were setting a considerable challenge, which will require research and consultation. It is essential that quality adjustments should command widespread support, and we recommended that a relatively high threshold be set for the introduction of the measures into the National Accounts.

This is related to principle C, where I believe that the Report raises a question rather than provide a definite answer. The issue can be explained by asking what would have happened if the existing direct output measures had been introduced for education forty years ago (in 1963). The output series, even with the quality adjustment, would largely have followed the number of school pupils. This number has not grown in line with real GDP, whereas inputs have grown at much the same rate as GDP. So that there would have been a long period when measured year-on-year output growth was below the growth of inputs. The government output indicators are largely demographically driven. Yet, as we note in the Report, the value of the output per person is rising on account of the growth of the real economy. The future earnings of the school pupils are rising on average by some 1½% per year. There is a complementarity between the provision of educational services and the increasingly productive private economy.

How should we respond to this? We say in the report that “account should be taken” of this complementarity, but this does not explain how. Several people have quite reasonably asked us to elaborate. One of them is Martin Weale, Director of the National Institute Economic and Social Research, who has suggested that the increased earnings are a relative price change. As he says, it is like a terms of trade effect, and, as such should not enter the measure of GDP. A country may be producing the same number of barrels of oil, and that is what should enter GDP, regardless of whether the relative price of oil has risen. The parallel that he has drawn with international trade is, I feel, a useful one, and a guide to how we should proceed. In the international trade case, the GDP calculation is supplemented by a real Gross Domestic Income (GDI) calculation that takes account of the increased purchasing power of exports. Following this parallel, we may wish to supplement the largely demographic output volume measure by an additional calculation of the “consumption equivalent” of the public output. As we stress in Chapter 1 of the Report, “no one single number will serve all purposes”.

This is not, however, the only response. The terms of trade interpretation attributes all of the quality change to the price element. But, as the late Sir Jack Hibbert emphasised to us, at the heart of the problem of measuring public sector output is the fact that we have no real guide to separating price and quantity. In certain cases, as we discuss in the Report, a quality improvement in the public sector is equivalent to getting a larger package. Robert Solow once gave as an example of productivity increase the case of a milkman placing 2 bottles on the step rather than 1. Does not the same apply to a fireman who is protecting 2 cars in my garage rather than 1? In the case of the fire service, a repackaging treatment of quality change, taking account of the increased property protected, may be more appropriate than a terms of trade treatment.

I turn now to the principles governing the measurement of inputs and productivity, beginning with Principle I. Early on in the Review, I encountered one aspect of ONS practice that seemed to me bizarre. ONS had since 1998 published volume indicators for government outputs and volume

indicators for government inputs, and these were “National statistics”; they also gave figures for productivity, obtained by dividing output by input, but it was explicitly stated that this was *not* a “National Statistic”. How applying the laws of arithmetic to two national statistics did not produce a national statistic seemed puzzling. However, behind this lay an implicit view of the margins of error. I believe that this underlines our call for there to be explicit margins of error, rather than the present implicit practice of treating statistics as either “good enough” or “not good enough” to qualify as a National Statistic. The old Central Statistical Office classification A, B and C had much to recommend it; a classification as “A “ corresponded to margins of $\pm 3\%$, B to $\pm 3\%$ to $\pm 10\%$, and C to more than $\pm 10\%$. It may then well be that dividing two A series gives a B series.

More generally, we urged that ONS examine the implied productivity series and apply “triangulation” (Principle H). Productivity change should be interpreted in the light of a range of other information. As we say, the triangulation principle may be implemented to different degrees. At the minimum, it involves checking for coherence with other evidence. This will almost certainly involve working closely with the relevant government departments. Now of course such discussions may be regarded as evidence of political influence, but it seems to me that a fair test is to ask whether the same discussion would take place with the British Cement Association if it were found that productivity in the cement industry had fallen, or with the Retail Consortium if productivity figures for retailing appeared out of line with expectations.

But the triangulation could be more formalised. It could involve systematic comparisons with departmental performance measures. Here of course it is clear that some departmental performance measures are concerned with total outcomes, such as reducing deaths from cancer, whereas the national income measure is concerned with the incremental contribution of government activities. The most ambitious level would be a government productivity measurement programme, such as that which, rather remarkably, was operated in the US from 1973 to 1994, covering two-thirds of federal civilian employees.

5. CONCLUSION

The Review was part of a dynamic process of improving the national accounts treatment of government output, which began many years before and which continues today. The issues raised have, in a number of cases, been resolved, while new issues are coming on the horizon. The international situation is rapidly changing, as EU Member States seek to meet the deadline of the 2006 National accounts. The UN System of National Accounts is again being revised. In this process, an important role is played both by official statisticians and by academic researchers. Acceptance by the wider research community is essential if new methods are to enjoy confidence.

A final implication is that increased statistical resources need to be allocated to measuring government outputs and inputs. With the benefit of hindsight, one can see that the move in the UK towards direct measures of government output increased the demands on ONS staff and departmental resources to an extent that was not appreciated. The public sector is an important part of the economy and resources should be allocated commensurate with its macro-economic significance.

References

Atkinson, A. B., (2005) Measurement of Government Output and Productivity for the National Accounts, Palgrave Macmillan, London.

Eurostat (2001) Handbook of Price and Volume Measures, Eurostat, Luxembourg.

United Nations (1993) System of National Accounts, United Nations, New York.

APPENDIX

Table 1: ONS Programme of Introducing Direct Output Measures

Function	Spending in 2000	When introduced	Main components of existing output indicator
Health	31%	1998	DH Cost-weighted activity index
Education	17%	1998	Pupil numbers with 0.25% quality adjustment
Social protection - social security administration	3%	1998	Number of benefit claims for 8 largest benefits
Public order and safety - prisons, courts and probation	3%	2000	Number of prisoner nights, number of court cases and cost-weighted activity index for probation.
Public order and safety - Fire	1%	2001	Number of fires attended of different types, other special services.
Social protection - personal social services	8%	2001	Number of adults and children in care. Number of hours of home help.

Table 2 Principles for the Measurement of Government Output and Productivity

- A:** Government non-market output should, as far as possible, be measured following a procedure parallel to that adopted for market output.
- B:** Output should in principle be measured in a way that is adjusted for quality.
- C:** Account should be taken of the complementarity between public and private output.
- D:** Formal criteria should be established for extending direct output measures to further services.
- E:** Measures should cover the whole of the UK.
- F:** Measures of inputs should be as comprehensive as possible, and should include capital services.
- G:** Criteria should be established for pay and price deflators applied to input spending series.
- H:** Independent corroborative evidence should be sought on government productivity, as part of a process of “triangulation”.
- I:** Explicit reference should be made to the margins of error surrounding national accounts estimates.

Discussion: Dr. Nick Sofroniou:

I would like to thank Sir Atkinson for his interesting paper. My comments and question concern the proposed education output measure involving the addition of a constant (0.25) times a measure of quality, examples of the latter being school examination results, pupil absenteeism and teacher-student ratio.

Some kind of 'value added' measure of institutional performance, looking at achievement gains conditional on student intake, might more directly reflect the educational period covered by each type of institution. However, comparing public examination results over time can be problematic given changes in curricula, marking standards and the wider cultural contexts in which students learn. Even the most complex item-response methodologies for educational testing are not immune from phenomena such as item-drift. Similarly, changes in pupil absenteeism over time may be more a reflection of wider social issues than simply school performance. The presented example of the use of the education output measure reconstructed for data going back to the early 1960s leads to concerns about the comparability of the measure over such a long period of time. My question is simply why a problematic quality measure was incorporated for the education output measure, when none was used with the other proposed output indices?

Dr. Nick Sofroniou
Educational Research Centre
Saint Patrick's College
Drumcondra
Dublin 9