

4 Methodology for the Future: The Principles

- 4.1** This chapter sets out a framework for considering the principles to be applied in developing the present measures of the change over time in government output at constant prices. The international guidelines described in Chapter 3 provide a point of departure, but a great deal remains to be done before the intentions signalled in these documents can be fully realised. The United Kingdom is in the vanguard of countries seeking to implement the European Commission Decision, and in some cases is attempting to break new ground. It is not therefore surprising that there has been considerable public discussion of the methods employed and their implications, not least on account of the political salience of public spending.
- 4.2** In view of this, we feel that it would be valuable to enunciate a set of principles on which to base the measurement of the output from government spending. These principles build on the current practice, often serving only to make explicit the implicit methodological approach, but a clear statement may inform the professional and public debate. The main principles we recommend are set out below, and cover the direct measurement of output, the measurement of inputs, and the measurement of productivity.
- 4.3** **Recommendation 4.1:** The direct measurement of the output from government spending, and the measurement of inputs and productivity, should be based on a set of principles, within the framework set by international guidelines.
- 4.4** The principles, which follow, are similar to those set out in the Interim Report, since most of them met with general agreement in the comments that we received. We have, however, reworded a number of the principles to clarify and emphasise the key issues. We have, in particular, thought more deeply about the initial Principle A, regarding the objective of output measurement.

- 4.5 At this point, we should stress that our concern is with government non-marketed output. This has two implications which underline our recommendations but which have provoked helpful comments on our Interim Report. First, we are dealing with activities that lie outside the market boundary that has been traditionally set for measuring national income. It can be argued that the concept of GDP should not be extended in this way. However, not only is such an extension entailed by the 1993 SNA and by our terms of reference but it also has intrinsic merits in terms of incorporating government activities into national accounts. Secondly, we are not examining the case for including other non-market activity. We are solely concerned with government output, not for example with household production activities. We appreciate that government services may complement, or substitute for, unpaid work in the home; the treatment of non-market services is being considered by the Panel to Study the Design of Nonmarket Accounts set up by the Committee on National Statistics of the National Research Council of the National Academies in the United States (National Research Council 2003). But our terms of reference are limited to government output.

Parallel with the Private Sector

- 4.6 The thrust of the SNA 1993 was, as emphasised by Neuburger and Caplan (1998), ‘to treat, as far as possible, public output in the same way as private output: the same general procedure can be used in both the public and the private sector’ (*Economic Trends*, 1998, p 31). This seems clearly right. The issues of measuring output and productivity apply across national accounts as a whole, and the principles applied to their measurement should, as far as feasible, be the same. This is particularly important in view of the transfers of activity that have taken place across the private/public boundary. It is evidently desirable that the relocation of an activity does not in itself lead to a change in the estimate of national output. Our terms of reference identify the need for comparability with measures of private sector services output and costs.
- 4.7 Therefore, we start from **Principle A: the measurement of government non-market output should, as far as possible, follow a procedure parallel to that adopted in national accounts for market output.**
- 4.8 Once we consider the public sector as supplying services, either individual or collective, it is therefore reasonable to ask: how far we can simply borrow from private sector experience, with appropriate modification? Can we learn from the private sector parallel how to implement the activity/output/outcome distinction drawn in the Eurostat Handbook? In the private sector, the concept of output with which we are concerned is the Gross Value Added: i.e. gross output less the goods and services bought in from other producers (intermediate consumption). Estimates of intermediate consumption are not available in time for the short term measure of the National Accounts and so early estimates of the chained volume measure of GDP (real GDP) are compiled using proxies for Gross Value Added. The ONS preferred proxy is turnover (output), deflated using a relevant price index. This assumes that the ratio of intermediate consumption to total output is constant, in chained volume terms, in the short term. The assumption does not always hold, for example if there were major changes to greater outsourcing of parts of production such as facilities management or IT support.

- 4.9** The proxy method just described has no counterpart in the case of the non-market output by the public sector, since there are no sales, but two other methods do have relevance. The first is to use change in the number of employees to represent changes in output. This input method is also currently used for a small number of private institutions, for example private hospitals. Where employment is used in the private sector, arbitrary allowances for changes in labour productivity in these private industries have been made. This approach is currently being reviewed as part of an industry by industry review of the service sector (Drew, *Economic Trends*, 2003).
- 4.10** The second method is to take direct output measures. For example, in the case of post and courier services, there are 14 indicators, such as the number of first class letters, tonnes of overseas letters, and use of courier services, which are combined with appropriate weights. This is a good starting point, but it should be noted that, according to the Eurostat Handbook section on post and courier services, ‘volume indicator methods based on detailed indicators of the many types of services provided, for example number of letters/parcels broken down by different postage rate, are B methods’ (paragraph 4.8.3.1). In order to achieve an A grading, it is necessary to take account of changes in quality.
- 4.11** Differentiation of the service is the standard way in which national accounts have taken quality change into account. In the postal case, we could say that there is differentiation of service, with the recorded delivery offering a higher quality than the standard service, and courier firms offering a different service level altogether. If more post is sent by recorded delivery, or courier, then the switch to a more highly valued service will lead to a recorded increase in output. In the case of transport, TP Hill suggests that ‘the amount of services produced must be based primarily on the number of passengers transported and the distance over which they are transported’ (1977, p 323), but that different qualities of service may be distinguished to take account of comfort, speed, punctuality and safety (Hill, 1977, p 323). On this basis, a punctual bus service is regarded as a different kind of service from an unreliable one. A bus company that becomes more punctual is then seen as selling more of a higher valued product. It may be argued that there is a difference between marketed and non-marketed output in that, in the market case, competition eliminates low-quality suppliers. But this assumes that competition works instantaneously, and means that no account is taken of the quality benefits from increased competition. There is, in this respect, a difference only of degree between the marketed and non-marketed sectors.
- 4.12** If a sufficiently detailed measurement can be conducted that the individual categories can be regarded as homogeneous, then a quality change shows up in the shifting proportions of the different services. As it is put in the Eurostat Handbook, ‘part of the quality change (that part due to compositional changes in an aggregate) can be captured by differentiating as many qualities of a product as possible. These different qualities are then in fact treated as different products’ (paragraph 2.4.3). But it is important to note the word ‘part’. Quality adjustment cannot be limited to differentiation of services.

- 4.13** If we consider a single service, the introduction of quality considerations raises questions concerning the definition of the service. Should quality enter the definition of the quantity unit? In the case of postal services, should the indicator be the number of letters posted or the number of letters successfully delivered? The degree of success in this case is largely under the control of the supplier, so that we may conclude that quality can be incorporated by taking as the unit the letter delivered, and it is for this that the consumer is paying. A reduction in lost letters then shows up directly as a quality gain. But how far should we take this? In measuring the output of a driving school, should we count only those lessons purchased that lead to success in passing the driving test? Or, since the school does not directly control the pass rate, should we count the number of lessons purchased, regardless of success in passing the driving test? One answer can be given in terms of the degree of *attribution*. If we are confident that the outcome is largely attributable to the supplier, then we may incorporate the quality element into the definition of the quantity. An example is where we treat quality change as ‘simple repackaging’: a quality improvement is assumed to be equivalent to getting a larger package. Higher-grade petrol gives you 10 per cent more miles to the litre, so it is equivalent to 1.1 units of the lower grade.
- 4.14** In reality, the situation is typically more complicated, since there are multiple attributes and we may be less confident about the attribution. Where there are multiple attributes these have to be weighted. For example for letters, one attribute is the percentage delivered at all to the right address, and a different attribute is the proportion of letters delivered by a certain time the next day. These different dimensions of quality need to be combined into an overall quality adjustment. In the same way, where we are less confident, or attribute only a fraction of the outcome to the supplier, then we may take the service supplied (number of driving lessons), rather than the service delivered, but make an overall quality adjustment based on expert evidence. Finally, if we are very uncertain of the relationship, we may decide that no quality adjustment is possible.
- 4.15** These examples from the market sector bring out the problems that arise in seeking to define output and show that it is hard to ignore the quality of the outcome. Outcome either enters in the differentiation of the service (recorded delivery) or in the unit measured (letters delivered), or in an overall quality adjustment.
- 4.16** A further issue illustrated by the postal services example is that of variation in cost of supply. Delivering letters to a terrace of houses opening directly onto the street takes less time than when there are long suburban front gardens. Should this be taken into account? On one view, we would not regard these as different services: the delivery of a letter is the unit of service regardless of its cost of production. If more people live in the suburbs, this does not show up as an increase in postal output. At the same time, the higher costs of transporting heavier letters are reflected in the differential tariff, and this is accepted as a differentiation of service. If people write heavier letters, then this is recorded as an increase in output. How does one distinguish these two cases? In the market sector, it can be argued that the producer makes the distinction in deciding on the price list, but it remains the case that the statistician has to decide how to group services. The degree of differentiation is a matter of statistical convention, not just market-determined.

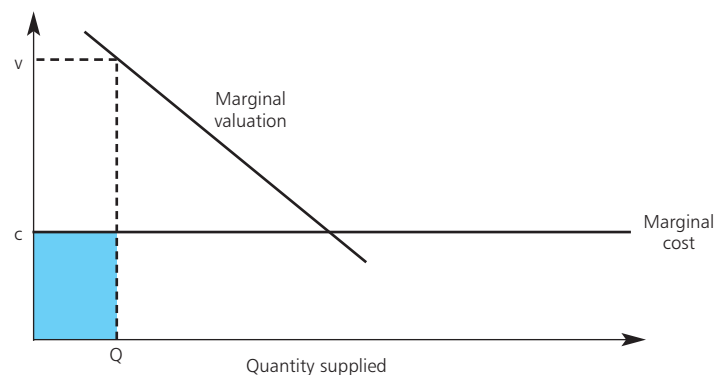
4.17 A final issue illustrated by the postal service is the ‘option value’ attached to the existence of a service. There is a value to a person even if he or she does not actually make use of the service. The notion is most commonly applied to natural resources, but the same applies to transactions closer to the market. A person derives reassurance from the fact that there is a local plumber, whose services could be secured in the event of a water leak. How is this recorded? If the person takes out central heating repair insurance, securing an option on the services of a plumber, then this market transaction reflects the *ex ante* value. Alternatively, the person pays *ex post*. Clearly, the year-to-year pattern will be different, reflecting the number of days of frost in the year.

The Concept of Added Value in Public Services: Individual Services

4.18 As we explained in Chapter 1, national income is both a measure of the major economic flows and an indicator of the contribution of economic activity to increasing welfare. GDP is widely used ‘by analysts, politicians, the press, the business community and the public at large as summary, global indicators of economic activity and welfare’ (SNA 1993, paragraph 1.68). In the case of final consumption by households, the justification for the welfare interpretation is that consumers are assumed to purchase an item until its marginal value, measured in terms of money, is equal to the price. If X denotes the quantity of the good, and p denotes the price per unit, then the value to the consumer of the marginal unit is p , and the total value is found by valuing all units at p , so the contribution to national income is pX . Goods for which there is a high marginal willingness to pay receive a larger weight in national income than goods for which there is a low marginal willingness to pay. It should be stressed that the resulting measure is not the same as consumer surplus; as noted in paragraph 1.22, we are not taking any account of the fact that the consumer would be willing to pay much more for the first unit of the good. The key element is that the welfare justification lies in measuring the added value to consumers. This value is inferred from the fact that economic agents are undertaking the transactions.

4.19 The problem in the case of non-market output is that there is no transaction from which the price or quantity can be observed. There are, in fact, two difficulties. First, there is no revealed preference by consumers, but, second, neither can the costs of supplying a marginal unit be taken as a measure of the individual or collective benefit. There is no reason to suppose that government output is supplied to the point where the benefit from a marginal unit is equal to the marginal cost of supply. This is illustrated in Figure 4.1, which shows a declining marginal valuation as more output is supplied, and, to simplify, a constant marginal cost of production. The total cost is found, in this case, by multiplying the marginal cost by the quantity, giving the shaded area. As we have seen, the convention that (output=input) is no longer one that we can regard as acceptable. So we cannot simply take the shaded area. What we have to attempt therefore is to measure the whole rectangle under the marginal valuation curve, i.e. vq . Even if we can observe q , the quantity supplied, we still need to construct, or find ways of inferring the marginal valuation, v , of that quantity supplied. This is particularly difficult in the case of those services where the nature of the service is not adequately defined, i.e. where there are no terms of sale specifying, at least in part, what is constituted by the transaction.

Figure 4.1



4.20 In order to consider the implementation of this approach with respect to individual services, we turn to the input/activity/output/outcome distinctions made in the Eurostat Handbook. To remind the reader, and taking the health service as an example, we identify the *inputs* as the time of medical and non-medical staff, the drugs, electricity and other inputs purchased, and the capital services from the equipment and buildings used. These resources are used in primary care and hospital *activities*, such as a GP making an examination or the carrying out of a heart operation. These activities are designed to benefit the individual patient. To the extent that they do, the health care provided constitutes the *output* associated with these input activities. Finally there is the health *outcome*, which may depend on a number of factors apart from the output of health care, such as whether or not the person gives up smoking.

- 4.21** Inputs are not an appropriate measure, for reasons already made clear, and, while activities may be the only available indicator and hence have to be used, they, too, are an intermediate variable. The relation between output and outcome, on the other hand, is less obvious, and encounters the problem of defining the quantity unit for the measurement of output that we have discussed above in the case of private services.
- 4.22** The first point to clarify is the difference between measuring total outcomes and measuring the contribution to outcomes of certain activities. A common objection to the use of outcome is that the status of the population is affected by many factors other than public spending. Examination performance depends on the efforts and work done by pupils. Parents who devote more time to teaching their children increase the level of education of a society in a way that cannot be attributed to public spending. This objection is well based, as we explained in paragraph 1.23. But it does not mean that outcomes are irrelevant. What it does suggest is that what we want to measure is the *incremental impact on outcomes arising from the activities of the public sector*. In the case of Education, the objective should be to measure the improvement in individual educational outcomes attributable to the schools.
- 4.23** Ideally, therefore, we want to measure the incremental contribution of the publicly provided service. In practice, it may be very difficult to attribute the outcome to different sources. This appears to be the reason why Hill (1977) argued against using examination results as a measure of the amounts of services produced by educational services. At the same time, he wrote: ‘An educational service is ... the additional skill or knowledge imparted in a pupil directly as a result of the instruction provided by the teacher.’ (1977, p 323). This appears to go beyond measuring the number of hours for which pupils are taught (the driving school case). It may be contrasted with the *Eurostat Handbook* reference to output as ‘the quantity of teaching (that is, the transfer of knowledge, successfully or not)’ (*Eurostat Handbook*, paragraph 4.12). This latter definition sounds more like an activity measure. Certainly, it does not seem reasonable to treat time wasted in the classroom as an output, other than as serving a purely custodial function. To quote Hill, ‘If the pupil’s qualifications and ability are such that he is incapable of understanding and absorbing the teacher’s instruction, there can be no change in his condition as a result of the teacher’s activity and no service is produced in these circumstances. The activity of the teacher is wasted and cannot count as productive’ (1977, p 324).

4.24 Approaching the question another way, we are firmly of the view that measures of output growth should *in principle* take account of quality change. We appreciate that quality has many dimensions, and that some will prove elusive, but in principle quality changes, whether positive or negative, should be taken into account. This may be seen as no more than an application of Principle A, seeking to follow a parallel procedure to that used in the market economy. As it was put by Professor Rhind, Chair of the Statistics Commission, ‘we cannot talk about the productivity of the car industry and how it has changed over twenty years simply on the number of cars being produced. Cars ... have far more functions, are safer and all the rest of it than twenty years ago. ... some sort of parallel to that may well have to be built into assessments of public sector productivity.’ (House of Commons Treasury Sub-Committee, Minutes of Evidence, 8 September 2004, HC 1039-i, Session 2003-04.) While it may be seen as a corollary of Principle A, we feel that it is best made explicit: **Principle B: the output of the government sector should in principle be measured in a way that is adjusted for quality, taking account of the attributable incremental contribution of the service to the outcome.**

4.25 As we have seen in the case of private services (paragraphs 4.11-4.15), there are at least three different ways in which we can approach the measurement of quality in the national accounts.

First, we can differentiate the services, with the aim of arriving at categories that can be regarded as homogeneous. A quality change is then captured by changes in the proportions of different categories.

Secondly, we can define the volume measure in terms of the degree of success.

Thirdly, the volume measure may be based on the level of activity, such as the number of lessons taught, but the contribution to outcomes introduced in the form of a quality adjustment. The volume measure would be ‘marked up or down’ by a percentage reflecting indicators of success and the contribution of the service to that success.

The Concept of Added Value in Public Services: Collective Services

4.26 The discussion so far refers to individual services: those consumed by individual households. The difficulties outlined above become more severe when we turn to collective services provided to society as a whole. Thorny issues are involved in devising direct output measures appropriate for COFOG categories such as General Public Services, Defence, Economic Affairs, and Environmental Protection. The Eurostat Handbook recognises the greater difficulties with collective services by accepting that either input measures or volume indicators of activity may be acceptable (under certain conditions) as B methods.

4.27 It is not always easy to draw a clear distinction between a collective and an individual service. There are spending functions that are largely individual in nature but which also contain elements of collective services. The same is true in reverse: some predominantly collective services have individual service elements. In both cases, we consider that these should be treated separately within the function. However, if new functions are to be covered, we feel that it should only be done after careful assessment and subject to a number of criteria, which we set out in paragraphs 4.40 to 4.44 below.

Complementarity between Public and Private Output

4.28 We begin by taking a long-run perspective, examining what would have happened if the United Kingdom had used direct measures to measure educational output since 1963. In Chapter 2, we noted that the CSO had used direct output measures in the 1950s and early 1960s, and that these showed that between 1959 and 1963, the direct method led to estimates of output increase that were significantly lower than those obtained using the (output=input) convention. The direct output measures were then abandoned in the United Kingdom until their use was revived following the UN SNA 1993.

4.29 The interpretation of national accounts over a 40-year period is open to debate. It can be argued that constant price series become increasingly unreliable if applied over long periods, even where appropriately linked. The separation of money GDP into price and volume components becomes increasingly difficult. At the end date, the nature of the economy looks less and less like that in the starting year. Goods and services have changed out of all recognition. In 1963 students did not have hand calculators, let alone personal computers. It is hard to construct robust price indicators. On the other side, there are those who accept the qualifications that surround the constant price adjustments but feel that the long-run perspective is a valuable one, and that the broad picture can be trusted. At the very least, taking a long-run perspective is a valuable thought experiment.

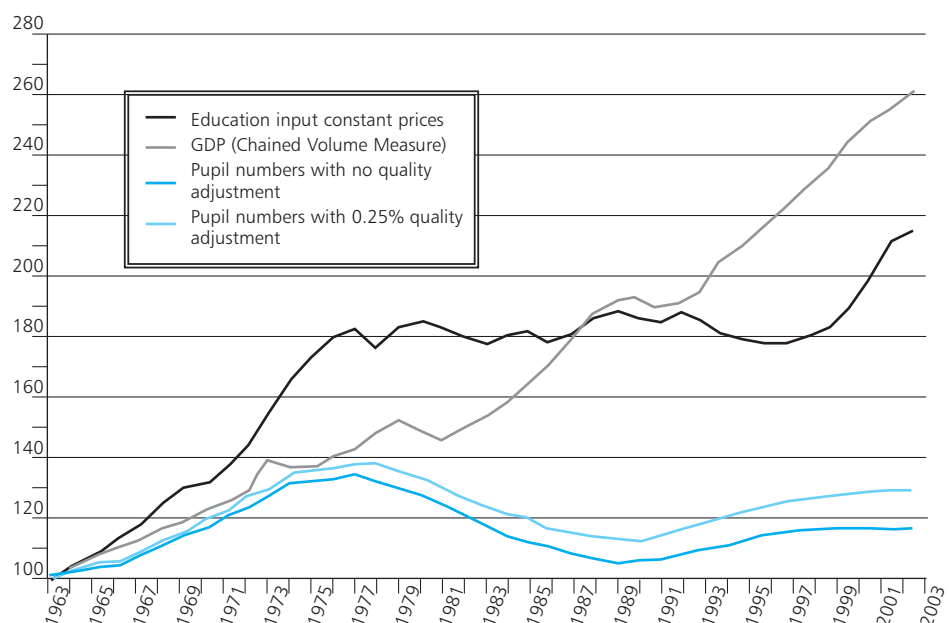
4.30 Figure 4.2 shows the implications of applying the current direct output measurement method to the education series over the period 1963 to 2003. Also shown for comparison is the deflated educational spending (1963 is the first year for which the Local Authority education constant price series exists; central government expenditure is a relatively small part of the total and is assumed to rise proportionately). Between 1963 and 2003, deflated spending rose by 115 per cent. Over the same period, GDP at constant prices is estimated (with the qualifications in the previous paragraph) to have grown by 162 per cent (shown by the grey line).

4.31 The starting point for the direct output calculation is the weighted number of full time equivalent pupils, and this is shown by the lowest series, marked by the darker pink line. The series does not include further or higher education, and the relevant spending is also excluded. The pupil numbers rose to reach a peak in 1977, 34 per cent above the 1963 figure, then fell, reaching a value in 1990 only 6 per cent above the 1963 level. The subsequent rise in the 1990s has led to a 2003 figure 17 per cent higher than 40 years earlier.

4.32 The darker pink line in Figure 4.2 shows the FTE pupil numbers adjusted by the quality factor of 0.25 per cent per year since 1963. Even with the quality adjustment, the growth in output is considerably less than the growth in deflated inputs. Taken together, they imply a substantial fall in productivity over the period as a whole. Between 1979 and 1989, for example, the implied productivity fall was some 19 per cent; in the previous decade it had been 17 per cent. For measured output to have grown at the same rate over the period 1963–2003 as deflated inputs, the quality adjustment would have had to have been some 1.5 per cent per year (geometric). For output to have retained its share of GDP, it would have had to have been 2 per cent. There is no reason to suppose that other periods would have led to very different conclusions. The baby boom may have caused a cycle, but in the long-run a demographic-based output measure is bound to show output of the service falling behind GDP where GDP per head is rising in real terms. As noted above, one has to be cautious in interpreting constant price series over a long run of years, but the effect is exhibited on a year to year basis.

4.33 There are several possible responses to this long-run picture. One response accepts that in a service like teaching there is little scope for productivity increase. As noted in Chapter 2 (paragraph 2.11), many public services involve an essential human input. There may be some scope for more efficient organisation, and for greater use of supporting equipment and staff, but there is no possibility of reducing the teacher/pupil ratio to the degree that labour productivity has increased in manufacturing.

Figure 4.2



4.34 The second response is to point to the complementarity between public and private output. To illustrate this, we begin with another example: the fire service. The present output measure is an ex post calculation of the actual number of primary and secondary fires attended (plus an allowance for fire prevention and special services). But this ‘incident’ count takes no account of the benefit derived from putting out a fire. Suppose that we consider the ex ante valuation placed on the existence of the fire service (the ‘option value’ – see paragraph 4.17). This valuation will reflect both the probability of loss and the magnitude of the loss. The direct output measure in effect takes account of the former but not the latter. This neglect of the extent of property protected does not seem reasonable. The ‘benefit’ of saving a house from a fire must have gone up with rising housing standards and increased household contents. (We have in mind here the replacement cost of the property, not the total cost that includes the value of the site.) The same applies to industrial and commercial property. The prison service provides a second example. The value to society may be seen in terms of the ex ante capacity to incarcerate people if the need arises. The value depends in part on the scale of property to be protected and of the economic activity that would be disrupted by criminal activity. The advent of Information Technology, for example, has both increased real GDP and increased the need for protection.

4.35 It should be stressed that we are talking here about the real value, not the increase due to inflation. Moreover, we are talking about the marginal valuation, not about consumer surplus. We are not seeking to attribute to the fire service the full benefit from its existence. We are not measuring the total area under the demand curve. Rather we are adding the marginal willingness to pay of different citizens. Two centuries ago, when the typical home contained little of value, the value of the fire service or of police protection or the courts in terms of domestic property was quite limited. Today, with homes resembling offices/factories, and with offices and factories full of expensive equipment, the value is many times higher. To this it may be responded that the typical person two centuries ago may have owned very little but he or she could much less afford the loss. This is true, but the fact that £1 meant so much more applies also to private output, to which we are adding public output to form GDP. The person could equally not afford to lose a week’s wages. Finally, we are not suggesting that the value of public sector output necessarily increases proportionately with the value of private sector output. If people chose to spend all of their increased real income on foreign travel, then there would be no increase in the value of home contents (although there would probably be more airports).

- 4.36** The point made here, is a quite general one: to a significant degree, the output of government services rises with the real value of private assets and incomes. We have just seen the force of this argument in the case of the protective services and it can be extended to education and to health care. Again the valuation of the output should reflect rising real output per head in the private sector. If we see the output of Education in terms of the acquisition of skills and qualifications, then their value increases with rising real earnings. If a university degree adds, say, 20 per cent to earnings, then today's degree adds 20 per cent of a larger number (even adjusted for inflation) than the degrees of a generation ago. It should be stressed that these considerations apply to *changes* in the valuation over time. It is the value of output relative to a generation ago that is affected. In practice, real earnings in the United Kingdom have risen by some 1.5 per cent per annum. Moreover, it is the general level of earnings that is relevant. There are clearly large earnings differentials between individuals, and between groups such as men and women, but it is the overall growth of earnings that we are seeking to capture.
- 4.37** It may be argued that education provided by schools is only partly responsible for the increased skills and qualifications. This is certainly true, but it is only relevant here if there has been a *change* in the extent the increase can be attributed to education. If 60 per cent, say, of earnings gain is attributable to school education, then in both years we are taking 60 per cent of a total earnings figure, which has increased, say, by 1.5 per cent per year. Secondly, it may be objected that we are double-counting the future earnings gain. The increased earnings appear both now, in the calculation of public output, and later when the worker receives the earnings. In this respect, however, the outlay is no different from any other purchase of an investment good. Finally, if the increase in the number of people with degrees leads to a fall in the earnings premium for graduates, then the value will not rise in line with average earnings. This is certainly true and needs to be taken into account.
- 4.38** In the case of health, rising real wage rates means that we attach a higher valuation to days lost through sickness absence. An extra week at work today is worth more than forty years ago. The same effect may apply more generally. The literature on Quality Adjusted Life Years has considered how the financial value to be attached should be adjusted over time. The answer given by Gravelle and Smith (2001) is that it should grow at approximately 1.5 per cent per year in real terms. Again, it is the overall change that is relevant. We are not seeking to differentiate between the values of extended life for people alive at a particular date.
- 4.39** In our view, this complementarity is important. Unless it is taken into account, there is a serious risk that the output of the public sector will be understated. This may be summarised in **Principle C: account should be taken of the complementarity between public and private output, allowing for the increased real value of public services in an economy with rising real GDP.**

Coverage of Output Indicators

- 4.40** The fact that national income is an indicator, not a precise addition over all the possible constituent parts, means that we should not expect an indicator of government output to be based on data that cover every single pound of government spending under a particular heading. At the same time, the international guidelines are right to stress the need for extensive coverage. The Eurostat Handbook requires for the A grading that the indicator should cover all services provided. Realistically, a balance has to be drawn between coverage and cost of implementation. We would expect that, within a government function, all major services should be covered. ONS practice has shown, for example in the new Health measures, that a high level of coverage can be achieved.
- 4.41** To make this more concrete, we believe that the procedure of defining direct output indicators within a government function should start by seeking to identify the services provided by government to households and firms, and attempts made to find data to reflect these services as comprehensively as possible, with appropriate allowance for quality change. The services should be the starting point, not the available indicators. If, initially, it is necessary to apply an indicator from another service, this should be explicit. A condition for the introduction of a new indicator should be that it covers adequately the full range of services for the functional area. The coverage of indicators within a function should be reassessed on a regular basis. Development of new output measures is discussed further in Chapter 6, but we would expect that the effects of its introduction had been tested service by service. A comparison would be made with the input-based estimates, and an examination made of the implied productivity estimate (see below).
- 4.42** We recognise that these criteria set a high standard. As in the existing functions covered in the United Kingdom, the initial estimates may not meet all these criteria, and there will need to be a continual process of revision and improvement. In order to meet the Eurostat deadlines, it may be necessary to introduce indicators before proper allowance can be made for quality change. As we have emphasised earlier, there are conceptual and practical problems to be overcome. But it is important that an initial evaluation be made of the proposed indicator, and that the shortcomings be clearly identified before the direct output measures are introduced, if only to warn users of the limitations of the output measures.
- 4.43** Once a new measure has been introduced, there should be provision for regular statistical review. Particular attention should be made to the consequences of changes in the machinery of government and in the methods of service delivery. The proposed method should be robust with respect to changes in the allocation of responsibilities between government departments. The method needs to be reviewed in the light of new developments, such as in the UK health care case the introduction of NHS Direct.

- 4.44** To sum up, we commend **Principle D: formal criteria should be set in place for the extension of direct output measurement to new functions of government. Specifically, the conditions for introducing a new directly measured output indicator should be that (i) it covers adequately the full range of services for that functional area, (ii) it makes appropriate allowance for quality change, (iii) the effects of its introduction have been tested service by service, (iv) the context in which it will be published has been fully assessed, in particular the implied productivity estimate, and (v) there should be provision for regular statistical review.**

Geographical Coverage

- 4.45** As was clear from the presentation in Table 2.1, the present direct output measures differ in their coverage across the United Kingdom. This is italicised in Table 4.1, where the marked entries indicate situations where the indicators fail to cover the whole of the United Kingdom. The reader of a volume entitled *United Kingdom National Accounts* might expect the coverage to be UK-wide, but this is not the case.
- 4.46** This lack of full geographic coverage is of concern because of the sums involved. It is particularly of concern given the long-standing differences between the constituent countries in services, methods of service delivery, and machinery of government. These differences have become even more important as a result of devolution. The output indicators should reflect the variation in services.
- 4.47** Moreover, the devolved governments, or their citizens, may well wish to have separate output estimates. These are of interest in their own right, allowing comparisons across countries. They also enter into the macroeconomic statistics for the different countries, such as growth rate of GDP. The same issues of comparability that have concerned Eurostat when dealing with EU Member States apply within the United Kingdom to the Devolved Administrations.
- 4.48** We have therefore been encouraged by the response of the authorities in the other three countries over the past year to the need for progress in meeting these challenges. We hope that they continue to give priority and adequate resourcing to this endeavour.
- 4.49** **Principle E: measures should cover the whole of the United Kingdom; where systems for public service delivery and/or data collection differ across the different countries of the United Kingdom, it is necessary to reflect this variation in the choice of indicators.**

Inputs

- 4.50** In Chapter 2, we stressed the need to consider not only the output side of the equation but also the input side. Correct measurement of inputs is necessary for several reasons. First, for most collective services inputs are likely to be the basis for output measurement for some time to come. Secondly, where there are direct measures of output, the deflated inputs provide a natural point of comparison, helping us to interpret the observed output changes. Thirdly, the deflated inputs have been used as the basis for statements about the implied productivity of the public services (see below).

Table 4.1

Function	Coverage	Main components of existing output indicator
Health	30.3%	<i>DH cost-weighted activity index</i>
Education	17.1%	<i>Pupil numbers with 0.25% quality adjustment</i>
Social Protection – Administration of Social Security	2.7%	Number of benefit claims for 12 largest benefits
Public Order and Safety – Prisons, Courts and Probation	3.0%	<i>Number of prisoner nights, number of court cases and cost-weighted activity index for probation</i>
Public Order and Safety – Fire	1.1%	Number of fires attended of different types, other special services
Social Protection – Personal Social Services	7.4%	<i>Number of adults and children in care. Number of hours of home help.</i>

4.51 We therefore commend two principles for the measurement of inputs. The first concerns coverage: the inputs taken into account should be as extensive as possible in their coverage. The inputs currently considered by ONS include all three of those listed above: capital, labour, and goods and services. However, the capital element is limited to capital consumption. Capital consumption refers to the depreciation of fixed capital, not to the opportunity cost of the capital being employed in the public sector, rather than in another, use. For any given type of asset, there is a flow of productive services from the cumulative stock of past investments. To illustrate, take the example of an office building. Service flows of an office building are the protection against rain, the comfort and storage services that the building provides to personnel during a given period.

4.52 We recommend that the appropriate measure of capital input for production and productivity analysis is the flow of *capital services* of an asset type. This involves adding to the capital consumption an interest charge, with an agreed interest rate, on the entire owned capital. It would be this total cost of the capital which would influence the leasing charge or the user charge under the Private Finance Initiative (PFI) where these alternative routes to securing capital services for the public sector were employed. Conceptually, capital services reflect a quantity, or physical concept, not to be confused with the value, or price concept of capital. Because flows of the quantity of capital services are not usually directly observable, they have to be approximated by assuming that the service flows are in proportion to the stock of assets.

- 4.53** The introduction of capital services is, as noted in Chapter 3, on the agenda for the revision of the SNA. This is a significant step. Given the capital-intensive nature of many government services, the inclusion of capital services ‘can give an entirely different picture of the costs of providing government services’ (Bos, 2003, p 122). In the present, more limited, context, it is not the initial levels that are affected but the rate of growth. Government services may always have been capital-intensive, so that their services remain relatively constant. If, for example, there has been a more rapid increase in the other inputs, notably labour, then the inclusion of capital services will mean that the measured increase in total inputs is lower than without the inclusion of capital services.
- 4.54** Capital services can be included in a variety of different ways, and the implementation of this recommendation will involve agreement on an appropriate convention. In considering such a convention, it will be useful to draw on the relation with the determination of the social discount rate to be applied in the ex ante appraisal of public projects (see for example Drèze and Stern, 1987, section 3.5.2). While it may be tempting to attribute a cost of capital related to the cost of government borrowing, it is not clear that this would be correct on either theoretical or practical grounds. The valuation of capital services should certainly not be affected by short-term changes in the sources of government financing.
- 4.55** As noted in Chapter 3, labour input may be measured directly in terms of the number of hours worked, with different skill categories being weighted. Alternatively, the labour input may be measured indirectly from total spending on payroll by deflating by an appropriate labour cost index. Ideally, both should yield comparable figures, but in practice approximations have to be made and the results may differ. We recommend that both methods be employed, and the findings compared and reconciled.
- 4.56** A significant element in the expansion of public spending in the United Kingdom in recent years has indeed been investment in human capital. Recruiting and training staff may appear as a recurrent cost but the benefits will only be seen later. This suggests that we consider the introduction of measures of human capital formation. The SNA argues that ‘while knowledge, skills and qualifications are clearly assets in a broad sense of the term, they cannot be equated with fixed assets as understood in the System.’ (SNA 1993, paragraph 1.52). It would, however, be possible and useful to create satellite accounts for human capital formation.
- 4.57** We therefore put forward **Principle F: the measurement of inputs should be as comprehensive as possible and in particular should include capital services; labour inputs should be compiled using both direct and indirect methods, compared and reconciled.**

Deflators

- 4.58** For each of the three main inputs (labour, goods and services, and capital), we need price indices appropriate to each of the spending functions. Each of these poses problems, as we illustrate for the first two.

- 4.59** In the case of goods and services, the requirements may be taken from the *Eurostat Handbook on Price and Volume Measures in National Accounts*. Specifically, to constitute an A method, it is required that (i) there be complete coverage, (ii) the prices should be purchasers' prices, (iii) quality change be taken into account, and (iv) there be consistency between the indicator and national accounts concepts. To these criteria, we would add the requirement that they should be sufficiently disaggregated to allow for changes in the input mix.
- 4.60** In the case of labour, the problems of indices for earnings are well known. Arguing by analogy with the Eurostat Handbook requirements for the prices of goods and services, we would expect that (i) there be complete coverage, (ii) the index should relate to actual earnings and not just wage settlements, (iii) the wage indicator should reflect total employment cost, including National Insurance and pension contributions, and (iv) that there be consistency between the indicator and national accounts concepts. There should also be sufficient disaggregation between types of labour to allow for changes in the input mix, eg a shift towards more highly skilled workers at higher earnings, or the reverse.
- 4.61** To summarise on pay and price deflators we propose **Principle G: criteria should be established for the quality of pay and price deflators to be applied to the input spending series; they should be sufficiently disaggregated to take account of changes in the mix of inputs and should reflect full and actual costs.** In the next chapter, we develop this principle by describing the criteria we believe should be applied.

Productivity and Triangulation

- 4.62** As noted in Chapter 2, the direct approach to the measurement of government output yields an *implied* measure of 'government productivity'. It is a residual. However, there is clearly a risk that the residual will behave in unexpected ways and that it will be dominated by the vagaries of the two measured variables. This implied measure may or may not be consistent with independent evidence on the productivity performance of the public sector. The national accounts necessarily reduce productivity measurement to a single number, and this aggregate statistic may need to be supplemented by richer information. In this section, we point to the need to obtain independent evidence on productivity, as part of a process of 'triangulation'. The independent evidence may be partial in its coverage or based on small-scale surveys, and would be assessed accordingly, but it should be part of the picture.
- 4.63** Starting from the equation $\text{Productivity} = \text{Output}/\text{Input}$, we would be seeking to combine evidence on all three elements, rather than assuming that productivity is simply the solution of this equation. This approach would be in the tradition of national accounts, where information from different sources has to be reconciled. This takes us back to the relation with microeconomic measures of public sector performance. As explained in Chapter 1, the national accounts measures do not have the same function. At the same time, they need to be coherent with the evidence from performance studies.

- 4.64** This leads us to propose **Principle H: independent corroborative evidence should be sought on government productivity, as part of a process of ‘triangulation’, recognising the limitations in reducing productivity to a single number.**
- 4.65** This process of triangulation can be conducted at different levels, involving different levels of statistical resources. How far it should be taken is therefore a decision that depends on the availability of resources. Here we distinguish three levels at which the issue could be approached.
- 4.66** As a *minimum*, the production process for the statistics of government output should involve a stage at which they are examined for coherence with other evidence. ‘Looking at the data’ should form part of the existing data production process, but this may need to be made more systematic. To draw an analogy with existing practice, if ONS finds that its statistics imply a sizeable reduction in the output of the oil and gas extraction industry, then this is discussed with industry representatives to see if the finding is consistent with industry experience. An explanation should be obtained for any divergence. The same applies to public sector output and productivity measurement.
- 4.67** In the present context, such cross-checking would involve consideration of the relation of the implied productivity estimates to departmental indicators of performance. As stressed above, these are not measuring the same thing, and there is no necessary reason why there should be agreement. It is, however, necessary that the reasons for any difference should be understood, particularly when the direction of change is different. It is possible for example that departmental performance indicators cover only part of the activities, and that resources have been transferred to ensure their fulfilment. The activities not recorded in the performance indicators may have seen a lower growth, causing an overall measure of output to show a lower rate of increase than the increase in overall inputs. Or it may be that the departmental targets are directed at quality improvements, and that these have been achieved but are not adequately captured in the output measure. The cross-checking may also reveal areas of the process that need further checking and refinement. The end product of such cross-checking would be, we envisage, published estimates of inputs, output and implied productivity that give due weight to the evidence and data regarding each – reconciled data, in fact, in line with long standing national accounting practice. It would be natural and helpful to accompany the published estimates by explanation. Equally, reference has been made to the relation between the indices of labour input obtained by deflating spending on pay and employment figures obtained directly from manpower or other sources. Such comparisons are undoubtedly made within ONS, but it would be helpful to the user if this kind of coherence check could be made more overt.

- 4.68** A *second level* would be an explicit attempt to relate the output and input indicators to departmental performance measures. From the side of ONS, this would mean a systematic examination of the relation between the direct output indicators and the PSA targets. In some cases, it will be clear that the PSA targets are concerned with total outcomes (such as reducing mortality rates from the major killer diseases), whereas the national income measure is concerned with the incremental contribution of government activities. These may well be moving in opposite directions for understandable reasons. From the side of departments, evidence should be assembled regarding productivity increase. Together, these actions should increase the degree of understanding of the relation between these two different sets of indicators.
- 4.69** The first and second levels could be encompassed within the regular ‘productivity articles’ that ONS has already decided to introduce. These articles can discuss measures of productivity in a particular sector of government output, help interpret the findings, and provide commentary on the underlying data issues. This commentary could have regard to the various relevant considerations discussed above. The articles could also bring together ‘triangulation’ evidence on productivity in the sector from different sources that may not be appropriate for the National Accounts but would help clarify the efficiency with which inputs are being used, in the context of the main policy and delivery priorities.
- 4.70** The Government is committed to improving efficiency in public services, following the Gershon Review whose conclusions were announced by the Chancellor in the Spending Review White Paper 2004. Departments have agreed Efficiency Technical Notes with the Treasury, setting out how they will achieve improvements in efficiency – the delivery of more or better services for the same inputs, or of the same service for reduced inputs. The methods of measurement proposed by departments will be of interest for ‘triangulation’: some examples are discussed in later chapters.
- 4.71** The *third, and most ambitious, level*, would be to initiate a government productivity measurement programme. This could draw on the US experience with the Federal Productivity Measurement Program conducted by the Bureau of Labor Statistics (BLS). This was brought to a close in 1994, but statistics produced covered a period as long as 27 years. The reasons cited for its closure were ‘budgetary constraints’ (Fisk and Forte, 1997, pp 19). This serves as a warning as to the resource costs, but does not indicate that the approach as such was a failure. As described in Chapter 3 (3.27-3.28), the programme used a large number of output indicators (more than 2,500 in 1994) to generate an output measure, disaggregated by functions, such as social services and benefits, or legal and judicial activities. But it would not be necessary to collect anything like this number of indicators to achieve a major improvement in the information available regarding public sector performance. In our view the design of such a programme for the United Kingdom deserves serious consideration.

4.72 On the input side, the programme collected indicators of the number of full-time equivalent employees (not adjusted for skill or experience) and of total employee compensation. From this, the BLS estimated indices for output per employee year (a 34.3 per cent increase between 1967 and 1994) and for unit labour cost. Two major limitations from our standpoint are that labour was the only input taken into consideration and that no allowance was made for changes in the skill composition of the labour force. It would however be possible, with sufficient resources, to overcome these shortcomings in a new government productivity measurement programme.

Margins of Error

4.73 Consideration of the national accounts estimates for government output, inputs and productivity raises for many users the question about the margins of error that surround these estimates. For 2003, the *Blue Book 2004* (Table 1.2) showed *individual* government final consumption expenditure of £140,870m, and *collective* government final consumption expenditure of £89,022m at current prices. How many of these figures are significant? Norbert Wiener, the physicist, may not have been wholly right to say that ‘economics is a one or two digit science’ (as reported by Morgenstern, 1963, p 116), but it might be more helpful to say that the latter figure is £89bn plus or minus (hypothetically) £2bn. On this basis the figure for 2003 would be clearly larger than that for 2002 (£81,835m), but the position would be unclear when comparing 1997 (£60,437m) with 1996 (£60,843m).

4.74 In the past, the Central Statistical Office gave such margins of error, in that it had an A, B and C classification, which corresponded to margins of error of ± 3 per cent, from ± 3 per cent to ± 10 per cent, and more than ± 10 per cent (Maurice, 1968, pp 40). These were described broadly as ‘good’, ‘fair’ and ‘poor’. These ratings were based on a subjective assessment. They were later dropped.

4.75 The current interest of ONS in this issue has been demonstrated by the project carried out under the auspices of Eurostat on ‘accuracy assessment of National Accounts statistics’. As the report of this project by Akritidis notes, ‘the quality of the statistical results is dependent not only in the quality of the underlying data, but also on the quality of the statistical process’ (2002, page 38). Part of the error is sampling error, where data are obtained by sample survey. Non-sampling errors are less easily quantified and can arise for a variety of reasons: incomplete coverage, failure to collect data, reporting error, processing error, and errors in making adjustments to the data collected. Moreover, the construction of national accounts involves combining data from many sources with complex interactions. Some elements are more reliable than others, but it may not be easy to determine their contribution to the resulting aggregates. It has also to be recognised that the reliability of estimates depends on the purpose for which they are being used. One user may be concerned with the level of GDP; another may be concerned with the growth rate over time. The first user might find a A/B/C classification of GDP sufficient; the second user would need to know about the correlation of errors in GDP over time.

4.76 We appreciate the difficulty of treating errors within the national accounts framework, but we would like to encourage ONS to give priority to work on accuracy assessment. We believe that users should be given as much information as possible about the margins of error surrounding the estimates: **Principle I: explicit reference should be made to the margins of error surrounding national accounts estimates.**

4.77 One issue that bears on the accuracy of the estimates of government output and inputs is the relation between the collection of statistical data and the auditing process. A statistical agency may rely more heavily on sampling, and may prefer more up-to-date, but less accurate, statistics to those that are more accurate but delayed. These differences are, however, ones of degree, and there are good reasons to suppose that a closer integration with the auditing process might lead to greater accuracy and reduce the reporting burden.

Summary of Principles

4.78 The principles are:

Principle A: the measurement of government non-market output should, as far as possible, follow a procedure parallel to that adopted in the national accounts for market output.

Principle B: the output of the government sector should in principle be measured in a way that is adjusted for quality, taking account of the attributable incremental contribution of the service to the outcome.

Principle C: account should be taken of the complementarity between public and private output, allowing for the increased real value of public services in an economy with rising real GDP.

Principle D: formal criteria should be set in place for the extension of direct output measurement to new functions of government. Specifically, the conditions for introducing a new directly measured output indicator should be that (i) it covers adequately the full range of services for that functional area, (ii) it makes appropriate allowance for quality change, (iii) the effects of its introduction have been tested service by service, (iv) the context in which it will be published has been fully assessed, in particular the implied productivity estimate, and (v) there should be provision for regular statistical review.

Principle E: measures should cover the whole of the United Kingdom; where systems for public service delivery and/or data collection differ across the different countries of the United Kingdom, it is necessary to reflect this variation in the choice of indicators.

Principle F: the measurement of inputs should be as comprehensive as possible, and in particular should include capital services; labour inputs should be compiled using both direct and indirect methods, compared and reconciled.

Principle G: criteria should be established for the quality of pay and price deflators to be applied to the input spending series; they should be sufficiently disaggregated to take account of changes in the mix of inputs and should reflect full and actual costs.

Principle H: independent corroborative evidence should be sought on government productivity, as part of a process of ‘triangulation’, recognising the limitations in reducing productivity to a single number.

Principle I: explicit reference should be made to the margins of error surrounding national accounts estimates.