Volume Title: New Developments in Productivity Measurement

Volume Author/Editor: John W. Kendrick and Beatrice N. Vaccara, eds.

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-43080-4

Volume URL: http://www.nber.org/books/kend80-1

Publication Date: 1980

Chapter Title: Current Efforts to Measure Productivity in the Public Sector: How Adequate for the National Accounts?

Chapter Author: Allan D. Searle, Charles Waite

Chapter URL: http://www.nber.org/chapters/c3916

Chapter pages in book: (p. 333 - 356)

5   Current Efforts to Measure Productivity in the Public Sector: How Adequate for the National Accounts?

Allan D. Searle and Charles A. Waite

5.1   Introduction

The federal government is currently engaged in a broad program to measure productivity in the federal sector.

The present effort to measure federal productivity is an immediate outgrowth of work undertaken in fiscal 1972 by a joint interagency task force (composed of the Office of Management and Budget, the General Accounting Office, and the Civil Service Commission) which collected data for fiscal years 1967–71 from 114 organizational units in 17 agencies representing about 55% of the federal civilian work force. In July 1973, the Office of Management and Budget endorsed the continuation of the project, and full responsibility for collecting input, output, and related information and for developing the productivity measures shifted to the Bureau of Labor Statistics (BLS).

In fiscal years 1974 and 1975, the Bureau expanded the data base, expanded coverage to include organizational units not previously covered, improved the quality of some of the input and output data, and refined the methodological procedures used by the original task force. Currently, about 65% of the federal civilian work force is covered by the BLS survey.

The question to which this paper is directed may be summarized thusly: Does this effort to measure productivity and the more modest efforts undertaken by certain state and local governments provide a basis for adjusting existing measures of government output in the national income and product accounts?

Charles A. Waite is at the Bureau of Economic Analysis. Allan D. Searle was a consultant to the Bureau from 1973 to 1978.

333

Our answer is No if government output in the national accounts is viewed as including the ultimate public goods and services which government provides, such as national security, education, or fire protection. However, the answer is Maybe if one views government output more narrowly, namely as the flow of governmental processes, such as passports issued, vouchers examined, or audits completed.

Although we prefer the broader concept of output, we find no consensus on this matter, and there is certainly no way to measure changes in output under this concept. Nevertheless, the narrower output concept also poses serious conceptual and measurement problems.

This paper will review alternative output concepts as well as pragmatic considerations which affect both national accounts and existing productivity measures. The current federal productivity measurement program is examined as to its possible use in improving the existing national accounts measures. Problems in developing productivity measures for the state and local government sector also will be examined briefly.

5.2   Government Output in National Accounts

Value added by government to national output, like value added in all nonbusiness sectors of the economy, is measured by total factor cost incurred. In the case of government, factor cost is confined to the compensation of employees. Interest payments are not regarded as measuring value added because they are subject to fluctuations which cannot be regarded as representing corresponding changes in the value of national production. (Of course, inclusion of business interest in GNP excludes explicit consideration of its behavior because any of its fluctuations not reflecting productive activity are offset by changes in profits.) Government output also excludes the return to government-owned property because the statistical basis for making a realistic imputation is inadequate. This is largely because a realistic market value of the rental value of government property is not available, nor is an estimate for the total value of government real capital assets (BEA is currently working to develop such estimates, however).

In real terms, government output is obtained for any year as the product of full-time equivalent employment in that year and base-year compensation per full-time equivalent employee.[1] We may add that the estimates are made in eleven categories including officers and enlisted

---

1. Subsequent to the preparation of this paper, the methodology for estimating real government output was revised as part of the comprehensive revision of the national income and product accounts. See the January 1976 issue of the *Survey of Current Business* (p. 22) for a summary of the revised methodology.

men by major military service, federal civilian workers, state and local education and noneducation employees. Also, federal and state and local work relief (and recently, public service job holders) are treated separately in the period they appear. Productivity changes within each category are conventionally assumed to be zero.

It is important to realize what is *not* measured. What is not measured is the amount of education, defense, etc., which the government provides from its expenditures. The government is not transformed from a final consumer to a producer selling products to individuals at an imputed value. This is because no one knows how to obtain an output index for any important government function except by measuring inputs.

This is not to say the present method of deflating government compensation cannot be refined. One possibility, which BEA is seriously investigating, is use of specification pricing for employment inputs as outlined by Denison. In specification pricing for any group of commodities or services, one seeks to select as specifications those characteristics which (1) can be readily identified and do not change over time so that one can get continuous series; (2) are reliably associated with the biggest price differentials—either because they are characteristically important to the buyer or because they are associated with such characteristics; and (3) are possessed in varying proportions by the class of commodities under consideration. One does not examine the use of the product or service. BEA is considering a classification of government compensation based on age, occupation, and education, all of which are associated with substantial earnings differentials. Base-year pay of age/occupation/education groups would be extrapolated by hours worked in order to obtain real compensation. The proposition is that work by government employees in the same age/occupation/education group with the same amount of hours worked represents a purchase of the same quality of labor at different dates. Statistically, BEA might use data on shifts in the pattern of the federal grade structure (adjusted for "grade-creep") as a proxy for shifts in age/occupation/education. The functional distribution of labor is irrelevant.

In summary, since BEA does not have market values to place on the services produced by government, it is not possible to price these services and therefore not possible to arrive at a real value of government output. As a result, BEA has adopted the convention of valuing government output in terms of input, namely employee compensation. Changes in real compensation/output result only from changes in employment. (Output in the household sector is derived in a similar manner.) Research is under way to modify this method by taking into consideration changing characteristics of the government work force that are associated with earnings differentials.

5.3   Alternative Concepts of Government Output

Many of the problems of measuring government output are similar to problems found in measuring output in the service sector of the private economy. Estimating the quantity and quality of service output is inherently difficult since no physical unit of standard quality is available. How does one quantify an auto repair in the private sector? Or police protection in the public sector?

At least two methods are used. They are ably explained by Ross and Burkhead in their excellent book, *Productivity in the Local Government Sector*. First, the number of direct outputs, for example cars repaired or arrests made, may be used. However, this does not take account of quality change, nor does the direct output represent the desired service output. For example, the number of spark plugs replaced may be a good direct output measure, but the service output is the properly operating auto. Similarly, the number of arrests made may be a direct output, but the service output is a decreasing crime rate. To overcome these problems, a second method may be used, that of calculating the effects or consequences of the service; in our example, this would be the number of properly operating autos, or the lower crime rate. This method has been criticized because it confuses services output and consequences of output. The consequences are not necessarily a direct result of the service, but of the service and the environment. In our example, your car can be inoperative because of bad roads and not because of too few spark plugs installed, or crime may increase because of the easy availability of handguns and not because of too few arrests made.

Difficult as private-sector service output is to measure, it has one distinct advantage over public-sector output: it has a market price, and thus the total value of private-sector services can be estimated.

Not only do we lack market values to place on services produced by the government, we also lack a consensus on what is really meant by public-sector output in the national accounts. It is clear how government value added is currently measured, but it is far from clear what concept of public-sector output is being estimated. Is the concept one of direct outputs, such as arrests made, tons of garbage collected, or bombs dropped, or is it more general, such as a secure environment or international peace? Our investigation turned up enough diversity of opinion on this point to make it difficult to assess the applicability of current productivity measurement efforts to the national accounts.

The statistical method employed in the national accounts for measuring government output, namely by measuring the value of inputs, is the most common, but perhaps the least sophisticated, method of measuring government output. Two other methods, each aimed at the different

output concepts discussed above, are used outside the national accounts. As we have seen, the first method estimates changes in output from changes in the quantity of direct outputs. The second method calculates output changes by estimates of changes in consequences; program-planning budgeting systems (PPBS) use this method. Bradford, Malt, and Oates (1969) have clearly drawn a distinction between the two. They separate output into what they term "D-output" (method 1) of the services directly produced, such as classroom hours taught, and "C-output" (method 2), the things of primary interest to the consumer, such as Johnny's ability to read—in other words, the consequences. As Ross and Burkhead (1971) pointed out, the C-output is functionally dependent on the D-output of that service, upon the D-output of any other public service that may influence it, and upon other environmental factors. To illustrate, let C-output be safety from crime and D-output be number of blocks patrolled by police. Although safety is related to number of blocks patrolled, many other variables also affect it. As is obvious, selecting C-output rather than D-output multiplies the problem of estimating public-sector productivity, particularly the problem of adjusting for quality change.

All studies of federal productivity, including the latest BLS effort, have focused on D-output. This is not only because it is more susceptible to measurement, but because it preserves the distinction between output and evaluation of effectiveness. Furthermore, to use consequences as estimates of government services output, as noted earlier, requires that one be able to separate those effects which directly result from the measured input from those resulting from the environment within which that service is performed. Moreover, changes in D-output measures are more nearly comparable to changes in most private-sector output measures. In the private sector, changes in output are measured in terms of physical units and not in terms of consequences, although consequences are a factor in determining relative values by buyers and sellers in the market place.

For these reasons, criticism of methods of measuring output in government productivity studies should not center on their use of direct outputs, but rather on other questions. Are the direct outputs final or intermediate? Should factor inputs in addition to labor be examined? Should the labor inputs be adjusted for characteristics associated with earnings differentials such as age, occupation, and education? Is it appropriate to derive a total government productivity index by combining the results for individual agencies? How appropriate is the sample of federal activities? Can the federal civilian results be extended to the military? To state and local governments?

5.4   Productivity Link between Input and Output

As implied earlier, if government direct output could be measured in physical terms and valued, then there would be no need to seek a suitable productivity measure with which to adjust the national accounts. Quantity, value, and price data would be available at the output end and presumably at the input end, and an index of value added at constant prices for the government sector would be attainable. The lack of comprehensive output quantity measures—under either the C-output or D-output definitions given above—precludes construction of both output-price and productivity measures for the government sector as a whole. Before casting aside all hope, however, some questions must be answered: Can deflated input values somehow be adjusted by some form of productivity measure to arrive at an *estimate* of government output? Or can the current employment series now used be similarly adjusted? If so, what is the appropriate form of productivity and output measure, and what are the assumptions involved? Are data available? What is now being done?

The link between input price and output price can be formulated in a variety of ways with output price as a function of productivity (or unit-labor requirements) and input prices. One formulation, for example, links input price, unit man-hours, unit materials requirements, unit labor requirements, wage rates, material prices, and output price.

Whatever use is made of government productivity measures requires certain compromises from an ideal concept and demands acceptance of a number of assumptions. If one accepts the national accounts assumptions that employment change is proportional to output change, then one might adopt suitable productivity measures as modifiers of the employment series.

Alternatively, one might apply a productivity index to the purchase price indexes to arrive at an estimate of output price. In this event, a productivity index using labor and materials as inputs would be required. However, on the assumption that unit material requirements tend to be stable in the short run and that unit profits data are inapplicable, one might be willing to confine attention to analyses of the type of labor productivity (output per man-hour) measure which would be an acceptable compromise.

For usefulness as adjustment factors in the national accounts, it would seem that the appropriate productivity measure would be one which consists of component unit-man-hour series weighted with labor cost—not with labor requirements (man-hours) as is usual in most of the Bureau of Labor Statistics industrial productivity series. The BLS

series are conceptually suitable for technological-change analysis. The national accounts, on the other hand, require dollar weights in order to attain consistency with the value and price series of the national accounts. It follows that the total man-hours indexes used to obtain the productivity measures would be weighted by the wage or salary rate in each job category.

These formulation problems are not particularly serious, however, for data are available to provide the necessary weighting systems for the productivity measures. Of paramount importance is another question altogether: How useful are the productivity measures which can be constructed for component parts (departments, offices, activities, etc.) of the government? The question inquires not only as to their accuracy, coverage, sample, representativeness, etc., but whether a conceptually meaningful combination of the various series can be made which would be of use in the national accounts system.

To illustrate the problem raised by the last question: If an improved method of constructing walls requires fewer nails and thus less labor time, yet carpenters continue hammering at a constant rate of nails-per-hour, productivity based on nails-per-hour is constant while productivity based on walls-constructed-per-hour is higher. Similarly, if better paint allows painters to use one instead of two coats per wall, and the number of gallons of paint applied per man-hour is constant, productivity from the painter's standpoint is unchanged, but productivity from the standpoint of the seller of the painted wall is doubled. If a "total" index of productivity for wall construction were constructed from the indexes for carpenters and painters (or carpenter shops and paint shops), the combined index would be unchanged. An index based on "final" output—(walls or houses) would increase because account has been taken of the change in relationship between the outputs of intermediate processes and of the final product. In addition to the problems involved in combining measures of productivity, the example highlights the importance of obtaining an appropriate measure of output.

Output measures for final product are not simply averages of component process indexes, but consists also of an interaction or "activity-mix" factor which can even pull the total index outside the range of the highest-lowest component index. A productivity index constructed as an average for component departments of an organization or of processes is likely to miss the improvement which comes from elimination or reduction in importance of tasks or processes or from the introduction of new processes. In general, this failure would tend to result in a downward-biased productivity measure.

5.5   Early Efforts to Measure Federal Productivity Change

Before examining the current federal productivity measurement program, it may be useful to review the relatively few earlier efforts by federal agencies to measure the productivity of an entire organization or selected organizational components, following broadly the review by Ross and Burkhead (1971). In 1932, the first empirical study of productivity in the federal government was done for the Post Office by Bowden. He identified thirteen separate outputs—seven classes of mail weighted by relative average labor time required to handle units of that class and six special-service transactions weighted by average costs (since labor cost is the predominant cost element in both categories, he argued that the two types of output were additive.) Bowden's inputs were full-time equivalent postal workers based on hours actually paid.

Bowden's research was not sophisticated by present-day standards (he omitted certain types of labor, did not correct for changes in hours worked, etc.) but his basic methodology has been used by all studies of federal government productivity since then.

That methodology may be summarized in four steps: First, the organizations to be studied within the federal government are identified. The choice depends on the availability of output measures. Second, the quantity of inputs associated with each output is estimated. Capital inputs are not measured. Third, each output is weighted by the percentage of total inputs, usually man-hours used in the production of that output in the base year. Last, both outputs and inputs are expressed as indexes; dividing the output index by the input index yields an index of productivity.

Other works in federal productivity measurement were published in the late 1950s by Vogely and by Litton. Vogely's primary effort was to devise output measures for the Bureau of Land Management of the Department of Interior. He divided the Bureau into its major programs —minerals, lands, grazing, and forestry—and identified output measures for each, such as cases closed. In the process, he developed unit cost measures which suggested the direction of productivity change.

Lytton estimated productivity indexes for five separate agencies, covering 56% of total federal nondefense agency employment and included 139 indicators of the work of the federal government. His measures of output were developed, in some cases, by the agency itself (as in the case of the Social Security Administration), but his list of outputs has led some analysts to question whether his outputs were final and whether double-counting existed. It is also difficult to ascertain how he combined the five indexes into one government-wide productivity index.

The first major effort at measuring federal productivity for more than one agency began in 1962 and was published in 1964 by the Bureau

of the Budget (Executive Office of the President, Bureau of the Budget 1964). The study covered 24% of total civilian federal employees in the fiscal year 1962. The period covered depended on the agency included and varied widely; as a result, no aggregate index was attempted.

The outputs selected met two criteria: (1) they were related to the mission of the agency as defined by law; (2) they were "final" in the sense that some person or unit other than the agency providing the service was the recipient. They were not adjusted for quality, although agencies submitted lists of quality improvements.

Some of the output measures used are comparable with the approach used in today's measures. For the Post Office Department, output measures were straightforward, being based on weighted series of twenty-one types of mail handled. For Treasury's Division of Disbursement, however, output was based on payments made and number of bonds issued. The output of the Department of Insurance of the Veterans Administration was based on policies in force (maintenance), number newly issued (separately by type), termination, and disability cases. For the Systems Maintenance Services of the Federal Aviation Agency, output was defined as the facility year, i.e., maintenance of an operating facility for a year; 318 such outputs were identified.

Three separate input indexes were used for each agency: (1) the unweighted man-hours; (2) dollar payroll costs; and (3) real total budget costs. For the latter, each agency developed its own price index to deflate current budget cost (how an agency could do this without making explicit productivity assumptions about its own work force is not explained).

The most important finding of the BOB study was that productivity could be measured at reasonable cost and that it could be a useful tool for government management and budgeting. The study found that the principal obstacle in the development of productivity measures was the requirement that products or service be measurable over time on a consistent basis. It is undoubtedly true that more or less satisfactory output measures were obtained for those agencies dealing with products and even well-defined services. Whether the output measures were "final" depended on the scope of the activity measured. How the BOB study might have dealt with agencies such as Defense, Labor's BLS, or Commerce's BEA, however, is not clear.

5.6 Current Productivity Measurement Efforts for the Federal Sector

In 1971, the federal government began a much more ambitious effort to measure federal productivity, initially mounted as a joint effort of the Civil Service Commission, Office of Management and Budget, and the General Accounting Office. In mid-1973, the responsibility for fur-

ther improvement, maintenance, and expansion of these productivity measures was placed in the Bureau of Labor Statistics. During fiscal 1972, a task force under the Joint Federal Management Improvement Program (JFMIP) surveyed the use of quantitative measures within seventeen Federal agencies and demonstrated how they could be improved for management purposes. Data were gathered for the construction of an overall productivity index. Special studies were carried out, aimed at improving measures of unit cost and effectiveness. The latest report shows that 48 agencies reported on activities of 245 organizational elements covering over 1.8 million staff years (Joint Financial Management Improvement Program 1975). For purposes of these measures, productivity is defined as output per unit of labor, rather than per unit of a combination of inputs. These measures compare actual changes in output per unit input, without regard to the mission of the organizational units. Thus, according to Ardolini and Hohenstein,

> they must be distinguished from measures of effectiveness which provide a means of determining whether an agency is proceeding toward its objectives by establishing a relationship between organizational actions and mission accomplishment. Both types of measures are useful tools: the productivity measures indicate changes in the real cost of producing an agency's output and the effectiveness measures determine the value of the agency's output to the recipients of its goods and services [Ardolini and Hohenstein 1974].

The authors continue with the statement that "ideally a productivity measure should relate final outputs to their associated direct and indirect inputs." "Ideally"—aye, there's the rub!

5.6.1 The Output Measure

The data-reporting instructions for this project remind the various agencies that "the output series for an organizational unit should ideally reflect every final output activity of the organizational element" and go on to say, "While it may be possible to separate output activities, assigning appropriate employee-years to each may be impossible. Therefore the output detail will be constrained by the availability of actual base-year, employee-year data or estimates of base-year weights (e.g., relative weights)." Cautions are included to count joint outputs only once. Also, the instructions state that each output measure should (1) consist of units which are relatively homogeneous with respect to their labor requirements; (2) should be repetitive; (3) should directly reflect the work loads of the organizational element (and not that of contractors, for example); (4) should reflect changes in output quality (e.g., adjust for quality change); (5) should reflect the amount of work

done in a fiscal year; and (6) should reflect the final products, services, and treatments of an organization. In connection with point (6), the instructions specifically call for the outputs to be final *"from the perspective of the organization providing the information."*

In fact, the output data are often "final" with respect to an organizational unit providing productivity data, but may be intermediate from an overall government point of view. For example, the productivity of payroll offices in issuing checks to employees is final from the viewpoint of the payroll office, but intermediate for the government as a whole.

At best, the total productivity measures are a weighted average of the separate productivity changes. Furthermore, not all the measures are output measures, even from this restricted view. For example, the JFMIP project identifies a number of tasks within the organizational element and defines output in terms of accomplishment of the task. Among the tasks are such activities as soil survey reports, contracts administered, patent disposals, weather observations, engine overhauls, claims adjudicated, number of applications examined, cubic feet of helium extracted, items printed. As one means of analyzing the quality of the data, the outputs were classified as (1) direct measures of work performed, e.g., engines overhauled; (2) partial measures, e.g., procurement contracts closed; (3) proxies, e.g., patients admitted; (4) population support, e.g., personnel served. According to the tabulations of the JFMIP, 85% of the outputs (weighted on a basis of man-years worked) are direct measures, with 5% more represented by partial indicators. Another classification shows that 71% of the outputs are "final," e.g., directly associated with performance of the mission, and the rest are intermediate.

Little need be said about the pitfalls attendant on the use of the proxy and population support measures. It is evident, for example, that a building maintenance measure based on square feet or floor area could result in a productivity measure which would show improvement if some of the work force were absent or if services were performed less frequently. "Population served" as measures for fire fighting and education are also unsatisfactory. But these categories of data represent only 10% of the total, and efforts are being made to upgrade the measures based on this type of data.

The 85–90% of the index which consists of direct and partial measures needs further scrutiny also. These series may very well be adequate for the specific tasks whose output they are designed to measure. The instructions do caution reporters to be alert for quality change and to adjust output measures in order to avoid measuring spurious productivity changes. Ways of measuring long-cycle production (e.g., ships) are presented and the need to match man-hours with output is stressed.

The man-hour measures include not only the direct labor but indirect labor which can be allocated to the output.

Many problems connected with measuring output trends arise because most federal activities result in production of services. The summary report for June 1973 (Joint Measurement Systems Project 1973) indicated that agencies sometimes state that an improvement in service has caused a productivity decline—that is, the measure takes no account of this type of quality change. The Bureau of Labor Statistics is attempting to correct this kind of misinterpretation of the instructions.

There are other types of quality change which affect the indexes. The BLS can solve some of these by constructing more detailed measures—for example, as BLS suggests, by means of separate indexes for high labor-intensive deliveries of mail to private residences and low labor-intensive deliveries to office buildings. Other quality changes—particularly those which cannot be solved by setting up such subproject categories—may be most difficult—for example, changes in time required from posting to delivery, or increased or decreased responsiveness of officials dealing with the public, or the efficiency of tax collection. Nevertheless, efforts need to be made. Sometimes an adjustment in quantity may be feasible on the basis of either the labor time or cost of an additional service feature instituted in connection with the basic service which is being measured.

5.6.2   Inputs

The early reports also state that two man-year indexes were prepared —one based on aggregate man-years and the other consisting of an index of current dollar compensation divided by an unpublished deflator for federal government general employee compensation covering salaries and fringe benefits. This latter index conforms more closely than the former to the concepts needed for national accounts deflators, but later JFMIP reports make no mention of it so far as we can ascertain. One hopes this index, as well as the associated productivity series, will be maintained together with subindexes as appropriate.

5.7   Productivity Measurement at the State and Local Level

As compared to the work done by the federal government, local efforts are fairly primitive. There is very little information available on the direct outputs of local governments; most do not keep data to measure outputs, and those that do have begun only recently. There is also controversy as to what is to be measured—direct outputs or consequences.

Hatry and Fisk, of the Urban Institute, in a report prepared for the National Commission on Productivity, state that "the main thing to be

said about the productivity of local governments in the United States is that little is known about it" (Hatry and Fisk 1971).

The earliest studies of local government output such as those of Ridley and Simon in 1938 or Schmardt and Stephens in 1960 define output in terms of consequences rather than direct outputs.[2] This emphasis was carried over into the area of planning-programming budgeting systems (PPBS) evaluation of government programs at the national and local level. However, the relationship of inputs to outputs was not examined in the early studies.

The first empirical studies relevant to local productivity and direct local output was published in 1969 by Bradford, Malt, and Oates, who estimated the increase in cost of direct output for local education, health and hospitals, police, fire, and public welfare activities. Their output measures, such as pupil-days for education and patient-days for hospitals, are crude, however, and their means of deflating input costs are open to question.

A more ambitious effort to examine local government productivity was undertaken in 1971 by the Urban Institute. The institute's position on output measurement is quite different from Bradford, Malt, and Oates in that it argues that output should be broadly defined to include effectiveness and quality, not merely efficiency and quantity. The implication is that multiple measures of local productivity should be developed. As an example, the institute suggests a way for estimating the components of the productivity of garbage collection. The output measure suggested is tons of garbage collected adjusted for quality change where the latter is calculated by rating average street cleanliness by a local official on a scale of 1 to 4 and the percentage of population expressing satisfaction with garbage collection based on a survey. The input measure is real dollar costs. The institute also examined ways in which police crime control output may be measured, but again focused on output as measured by consequences or effects rather than from services directly produced.

Hatry and Fisk could find no local government function for which nation-wide productivity data had been or could be calculated. They found no consensus, at the local level, of what to measure, much less how to measure it.

They did identify at least three types of statistics collected by local governments which might be used in measuring productivity—cost data, work load measures, and, in a very few cases, effectiveness measures. Cost data, such as dollars expended by function or program, are most common, but they are rarely expressed in real terms. Moreover, existing measures of price change for local governments, such as BEA's implicit

2. From Ross and Burkhead (1971).

deflator for state and local government, are not available by function or level of government. (However, BEA is currently working on detailed deflators and expects to publish them in the near future.)

Work load measures, such as miles of streets swept, number of students per school, etc., are collected by many governments. Hatry and Fisk found that the most ambitious programs were in Chicago, Dayton, Fort Worth, Los Angeles, New York, Philadelphia, San Diego, and Savannah. Relating costs to work loads is also done in some cities, but rarely in real terms.

The measurement of effectiveness, or the quality of the government service as it impacts on the citizens of the community, has been tried in only a few jurisdictions, such as Chicago and Dayton.

Information collected by the federal government bearing on the effectiveness of local programs is beset with many problems. Important areas of government activity, such as garbage collection and fire protection, are largely ignored. Reports are often voluntary and response rates poor. Output data are not related to input data. Definitions are inconsistent among governments. Most important, none of the special surveys or the routinely collected data compiled by federal agencies or others have been specifically designed to provide nationwide information on state or local government productivity.

In summary, although local and state governments comprise over two-thirds of government output as now estimated in the GNP accounts, efforts to measure their productivity are still in their infancy and lag well behind federal efforts. Until new data are collected and existing data improved, it will not be possible to produce productivity measures suitable for use in the state-local sector of the GNP accounts.

5.8   Do the Federal Productivity Measures Meet the Needs of National Accounts Measurement?

5.8.1   In Concept?

If one's concept of government output includes the consequences of government activity, such as national security, the new federal productivity measures are of little use for adjusting the existing national account measures. As noted, BLS does not attempt to estimate the real market value of government services. If such data were available, they could be used with presently available input data to derive real value added for government. With such an output measure, productivity measurement would be redundant. Instead, the current federal productivity measurement effort is studying the output of processes, for example, claims processed. As illustrated earlier, the weighted average of productivity changes in the processes which go into the provision of a final

product is not the same as the productivity gain in providing the final product. Eliminating, reorganizing, and replacing processes is the very essence of a most important way that productivity rises in the private economy. As noted, adjusting existing national-account government output measures by the new government productivity indexes would not give one the amount of defense, education, or other products the government provides. Neither could change in output per man-hour be used to adjust for change in labor quality.

5.8.2 As a Practical Alternative?

But this verdict cannot be the end of the matter, for we are a long way from measuring the market value of government services. A more reasonable approach must take realities into account and assess whether the use of currently available productivity measures would improve the national accounts. This paper is a modest effort to provide such an assessment.

The following section will touch on some of the more important measurement problems.

First, the new federal productivity measures combine intermediate with final outputs for individual agencies, although there is an attempt to avoid this. This duplication cannot be altogether offset in the weighting system, because of the problem of "activity-mix change" to which reference has been made earlier.

Second, the combining of agency outputs into a total ignores the interagency mix problem. The aggregate productivity measure represents a weighted average of agency measures.

These two points are related, but the current effort seems to view the *intraagency* duplication as an error to be corrected, but to view the *interagency* duplication as desirable. From the standpoint of the national accounts, both create problems, and we would suggest "netting out" of "final" *outputs* of one unit which are inputs to another, on an *intraagency* and *interagency* basis. Certainly, for use in national accounts, this netting would be essential. A year-to-year assessment would have to include an estimate of the bias caused by failure of the current indexes to measure the productivity change originating in "activity-mix" changes. It may be that in some years during which there has been little organizational change, the current measures would suffice. In other years, some estimates might be made—based perhaps on cost-effectiveness studies.

Third, many of the measures, as indicated above, are not really based on direct output or even intermediate output, but are in the nature of proxies, consisting of such measures as "personnel served" or "population support." We think these, together with indirect measures, may amount to 10–20% of the coverage, in weighted terms.

In order to ameliorate this problem, further experimentation might be conducted on measurement of final outputs of agencies where measurement problems are especially severe. For example, Marvin Mundel (1973) prepared an experimental measure for grants administration activities, in connection with the research work preceding the institution of the JFMIP project. He discussed two measures: (1) a performance (output per man-hour) measure constructed as the product of "standard time" for each current period, times quantity of work done, and divided by actual current man-hours; (2) a productivity index constructed by multiplying quantity, each year, by "standard time" of each type grant in the base period, summing, and dividing by actual man-hours. The first index could *decline* if today's actual performance compared to expected (standard time) is relatively poorer than yesterday's performance. The second index might increase if today's output per man-hour increased with respect to that of the base period. This latter concept seems more closely related to the productivity measures discussed earlier.

What makes Mundel's approach interesting, however, is not so much the choice offered between two indexes, each suitable for its own purpose, but his attempt to quantify *final* output of a service agency. In essence, he examines various kinds of grants (the end products) and then traces the associated man-hours and standard times throughout the agency.

Fourth, labor is treated as a homogeneous input. The experimental work which led to this study made mention of another labor series based on use of a deflator of current dollar compensation. This series, or some version of it, might be examined as an alternate labor measure. Also, such characteristics as wage rates, occupation, education, and experience might be considered along the general lines indicated by Denison and the BEA's experimental work along similar lines.

Fifth, capital inputs are not included. It is perhaps unfair to insist on this point in view of the dearth of work in this area, but the development of these measures would greatly enhance the results.

Sixth, it should be recognized that while the coverage of the federal civilian sector alone is relatively large, this coverage is not based on a random sample. One does not know how representative of the government's total civilian productivity experience the coverage is. This is because availability of output data happens to have been the most important criterion for inclusion. All agencies having 200 or more employees were asked to participate in the project. From this high point, fallouts occurred as organizations were unable to identify measurable outputs or matching inputs. As a result, the sample may not be as representative as one could wish, being biased, perhaps, by data from agencies with more routine or mechanized processes, more readily

organized data, or even more enthusiasm for the project. While the overall coverage is good, a more nearly representative sample should be sought. Coverage of some important government departments is well below average—for example, Defense (35%), State (9%), Commerce (23%)—compared with high coverage for Veterans Administration (95%), Transportation (85%), and HEW (79%).

5.8.3   As to Comprehensiveness?

The Federal productivity measurement project covered 65% of federal civilian employment in fiscal 1974, and coverage may eventually increase to 70%. The 1974 data include 1.8 million man-years out of a total of 2.8 million. While this coverage appears impressive, it is much less so when it is adjusted by removing government enterprises such as the Postal Service whose output is excluded by definition from general government in the GNP accounts. (Postal output is included with private-sector output in the GNP.) Excluding enterprises, the coverage drops to 50%. More important, it represents only about 25% of total federal general government (civilian and military) man-years, and only 8% of total general government (federal, state, and local) man-years.

In other words, even if the GNP accounts were to incorporate the new federal civilian productivity measures, it would have only a negligible impact on the current government output measures.[3] Certainly, even the most enthusiastic backer of the current productivity effort would not suggest using the new measures to estimate productivity in the military or the state-local sector.

5.9   What of the Future?

As discussed in the previous section, current federal productivity measures are less than ideal in concept and comprehensiveness, and have several statistical shortcomings from the standpoint of the national accounts.

However, this judgment should not be interpreted too harshly. First, there appears to be little agreement on the precise concept of government output, and the broadest output concept has never been calculated for the government as a whole. Second, although the existing measures cover only a small part of total government, they may point the way to new efforts covering the rest of the federal sector and the state and local sectors. Third, we believe many of the statistical problems in the BLS project can be overcome, perhaps by the means suggested earlier.

3. Rough calculations by BEA indicate that total government output in 1974 would be only 1% higher than currently estimated if the new federal productivity measures—available back to 1967—were used to adjust that part of federal output covered.

Nevertheless, it would be unwise, in our judgment, to introduce the existing federal measures into the national accounts. National accountants have to do more thinking about what should be measured as government output and how to measure government input, and productivity experts need to broaden the scope and refine the methods currently employed in measuring public-sector productivity. It is encouraging to see the BLS program underway, and we hope that eventually it will help to provide some of the tools necessary to construct a much-improved measure of government output.

References

Ardolini, Charles, and Hohenstein, Jeffrey. 1974. "Measuring productivity in the federal government." *Monthly Labor Review* 97 (Nov.): 13–20.

Bradford, D. F.; Malt, R. A.; and Oates, W. E. 1969. "The rising cost of local public services: Some evidence and reflections." *National Tax Journal* 22 (June): 185–202.

Hatry, H., and Fisk, D. M. 1971. *Improving productivity and productivity measurement in local governments*. The Urban Institute, Working Paper 2700–01.

Joint Financial Management Improvement Program. 1964. *Measuring productivity of federal government organizations*. Executive Office of the President, Office of Management and Budget.

————. *Productivity programs in the federal government, FY 1974*, vol. 1, *Current efforts and future prospects*.

Joint Measurement Systems Project. 1973. *Measuring and enhancing productivity in the federal government*, pt. 3, *Summary report*. Executive Office of the President, Office of Management and Budget.

Mundel, Marvin D. 1973. Measuring productivity of grants administration activities. *Special studies of measurement problems*, Special Report no. 3, section 2, Joint Measurement Systems Project, Executive Office of the President, Office of Management and Budget.

Ross, J. P., and Burkhead, J. 1974. *Productivity in the local government sector*. Lexington, Mass.: D. C. Heath.

Comment   Jerome A. Mark

The problems associated with measuring government output in the national accounts are perhaps the most frequently discussed and least resolved of all the problems with the accounts. But the need for some improvement in the existing procedures has long been recognized. In view of the recent attempts to measure productivity in the public sector, it is very appropriate that Searle and Waite examine these measurements to see how adequate they are for use in the accounts.

Their paper reviews the various conceptions of government output and the problems associated with implementing them. It describes how government output is currently being measured in the accounts and what form productivity adjustments should take to transform input to output. The paper then examines the productivity measurement work in the areas of federal, state, and local government to assess the usefulness of the results of these efforts for the accounts. Since productivity measurement at the state and local level is very limited, this assessment is almost entirely devoted to the current measurement program for the federal government.

This program to measure the productivity of federal agencies is a continuation of the work undertaken by a joint interagency task force composed of the Office of Management and Budget, the General Accounting Office, the Civil Service Commission, and the Bureau of Labor Statistics. It was established as a continuing program in July 1973, with the BLS given responsibility for collecting input, output, and related data and for improving the measures; and the Civil Service Commission, the General Services Administration, and the Joint Financial Management Improvement Program (consisting of the GAO and five executive departments) given responsibility for productivity enhancement efforts.

I mention the productivity enhancement aspect to point out that the main thrust of the program has been toward the development of productivity measures as an aid to government managers for decision-making, planning, and evaluation. Some functional groupings whose activities may be intermediate to the total government are useful for these purposes, whereas they would not be for national accounting purposes.

The federal sample includes 245 organizational elements from 48 agencies encompassing 65% of total federal civilian employment. It includes civilian components of the military, but does not include the armed forces directly. Government enterprises which, in the national accounts, are in the private sector, are also included in the sample. Their exclusion reduces the employment coverage in the federal general

Jerome A. Mark is with the Bureau of Labor Statistics.

government to about 50% and the number of organizational elements to 232. Separate labor productivity indexes were developed for each of these organizational elements and these in turn were aggregated into functional areas of government activity.

In determining the output indicators upon which the productivity indexes were based, the agencies, with some assistance from the BLS, had to identify specific units of service which were final to the organization, quantifiable, and fairly homogeneous over time. The nature of the indicators varies substantially; they include such items as currency notes delivered, trademarks disposed, passports issued, tanks repaired, weather observations, square feet of buildings cleaned, and deportable aliens located.

The question Searle and Waite initially address themselves to is, Does this effort to measure productivity provide a basis for adjusting existing measures of government accounts? Their answer is No, if government output in the national accounts is viewed as including ultimate public goods and services which the government provides, such as national security and education.

But Searle and Waite point out that, although this concept may be preferred (and there is no consensus on that), it cannot be implemented. Therefore, a narrower concept, the flow of government processes, must be considered. Under this concept, they say there may be possibilities for using the results of the efforts to measure government productivity.

In reviewing alternative concepts of government output, Searle and Waite point out that the problems are similar to measuring output in the private service sector, and thus are inherently difficult. Based on some work of Ross and Burkhead, Searle and Waite cite two methods for measuring output in service and/or government activities—either counting the number of direct outputs (such as cars repaired) or calculating the effects or consequences of the service. They feel the first method, counting direct output, does not take account of quality change, nor does it represent desired service output. The second method—calculating the consequences of services—is criticized because consequences are not necessarily a direct result of the service.

They point out that all studies of federal productivity, including the BLS one, have focused on the first method of counting direct output, which they feel is appropriate not only because it is more susceptible to measurement, but also because it preserves the distinction between output and an evaluation of effectiveness. Moreover, it is more compatible with private-service sector output measures which are in terms of physical units.

Their assessment of the methods of measuring output in government, therefore, is in terms not of the appropriateness of the concept which they accept as being direct output of specific government activities, but

rather of how adequate the measures are which have been developed. On this basis, they make six observations on the problems with the currently available productivity measures: (1) The measures combine intermediate with final outputs for individual agencies—although there is an attempt to avoid this. (2) The combining of agency outputs into a total ignores the interagency mix problem. (3) Many of the measures are not based on direct output, but are proxies. (4) Labor is treated as a homogeneous input. (5) Capital inputs are not included. (6) The coverage is not based on a random sample which, although large, may not be representative.

For these reasons, they conclude it would be unwise to introduce the existing federal measures into the national accounts.

The criticisms of the current federal government productivity efforts that Searle and Waite have made are in part valid and I agree with some of them, but I am more optimistic about the potential usefulness of the measures for the national accounts. My difference with the paper largely is in the perspective or framework in which the work is being assessed.

There *are* some serious limitations with the existing measures, and I would not recommend *blanket* inclusion of the current measures into the national accounts. In my opinion, it is not a question of the measures failing to meet an "ideal" definition of direct output; rather, it is whether better estimates of productivity change for component parts of the federal government sector can be derived from the measures developed than the assumed zero productivity change currently being used for the entire government sector.

Let me be somewhat more specific with regard to the paper's criticisms of the project's measures.

On the intermediate-final output problem there are, as Searle and Waite indicate, two aspects to it. The first relates to those output indicators for activities which are intermediate to the organizational element, whether or not the functions of the organization as a whole are external to the government. The second relates to those organizations whose output indicators are final to the organization, but whose organizational output is intermediate to other government agencies.

On the first—the intraagency problem—it may not be as serious as Searle and Waite imply. First of all, over 70% of the indicators presently used are final output indicators for the organizational elements, and the employee-years associated with those indicators represent almost 90% of the sample. Second, the problem is being reduced insofar as reporting is concerned. The reporting instructions do stress the need for final output measures, and BLS has worked closely with many of the agencies in examining the indicators and eliminating the intermediate measures when possible.

With regard to the interagency duplication, for national accounts purposes, of course, it would be essential to net out these intermediate outputs. For the purposes of the current program which are directed toward using the measures as an aid in the enhancement of government productivity, the combining of measures for organizations whose final outputs are intermediate to other government organizations is useful, particularly when examining functional areas across agency lines; for example, examining the productivity changes among support service organizations within the federal government.

The interagency duplication can be reduced since separate productivity indexes for the 245 organizational elements have been developed and efforts can be made to net out organizations with activities that are intermediate to the federal government. Organizational units providing support services such as all the elements in the General Services Administration, the Civil Service Commission, and the Defense Supply Agency are cases in point. Exploration of the possibilities along these lines would have been useful in the paper.

With regard to use of proxy indicators, which represent only about 10% of the coverage, it should only present a problem if a bias exists, and this hasn't been examined in this paper, but would have to be in any assessment.

On the treatment of labor as a homogeneous input, in order to separate the quality changes in labor input from productivity changes in the output-per-employee measures, it would be desirable to adjust the present measures for changes in the composition of labor input.

But for the purposes of using the measures in the national accounts, I do not believe the criticism is relevant. The purpose for using the productivity indexes from the federal project in the existing accounts is to adjust the employment change to derive an estimate of output change. For this purpose, it would *not* be desirable to make the adjustments for quality change in labor input to the productivity measure. If that were done, the resultant output would not reflect the direct effect of changes in labor quality on output. The unadjusted measures, as developed in the project, would be the appropriate ones.

If the Bureau of Economic Analysis introduces the adjustments to labor inputs to take account of shifts in the composition of labor along the lines Searle and Waite mention, then the project productivity indexes needed to transform government input into output would have to be modified, but until then, in this connection the present ones are the appropriate ones.

On the absence of capital inputs from the measures, I also do not believe that the criticism is relevant. This is only a limitation if we are concerned with developing a total factor productivity measure for the federal government. However, again as in the case of the labor quality

changes, it is not a limitation if the productivity measures of the project are used as an independent method for deriving government output. As in the previous case, if a total factor productivity measure could be developed and were used, the resultant output measure would not reflect the direct effects of capital inputs.

Finally, with regard to the absence of random sampling, I agree with Searle and Waite that there could be a problem of representativeness of a measure for the entire federal government. It would not arise, however, from an absence of random sampling since all agencies with 200 or more employees were asked to participate. If present, it would arise from the inability of some agencies to respond because of inadequate data. Of necessity, the availability of adequate output and input data had to be the principal criterion for inclusion.

In the eventuality that there might possibly be component measures useful for the accounts, they might, in addition to being included with their individual weight, also serve as a basis for imputation to uncovered parts. Imputations are used directly or indirectly in many components of the existing accounts. For example, a significant portion of the accounts rests on the use of wholesale price index information, and yet for a substantial part of the WPI imputations are used. Even in the part that is directly collected, coverage is limited and not based on a random sample.

In summary, then, in assessing the usefulness of the results of the current effort to measure productivity in the federal government, I believe they have to be examined in relation to what is currently being done—namely, assuming zero productivity change. This approach consists of addressing the following questions: Are the estimates of productivity derived based on reasonable procedures? Do they result in biased estimates? Can appropriate adjustments be made to overcome some of the problems, such as the intermediate output problem? How do the methods employed compare with some of the procedures and imputations currently used for deriving measures in the private sector of the accounts?

It is only on the basis of such considerations that the possible usefulness of the measures for the accounts can be assessed.

Searle and Waite point out that even if the accounts were to incorporate the federal general government productivity measures, it would only have a negligible impact on the total government output measures. I would agree that the potential effect is very small, and it certainly points out the need for additional, extensive efforts in the difficult area of state and local government productivity measurement. Perhaps, as Searle and Waite state, "although the existing measures cover only a small part of total government, they point the way to new efforts covering the rest of the federal sector and the state and local sector."

I share their belief that "many of the statistical problems in the BLS project can be overcome." Moreover, I agree that blanket inclusion on an overall basis of the productivity measures now available is not warranted. However, I do think that some degree of selective inclusion may be feasible and worth exploring, and with the continued development of the government productivity measurement program, more extensive inclusion might be possible.