Omega ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Omega

journal homepage: www.elsevier.com/locate/omega

Measuring effectiveness of production in the public sector ☆, ☆☆

Finn R. Førsund [a,b,*]

[a] Department of Economics, University of Oslo, Norway
[b] Norwegian Defence Research Establishment (FFI), Norway

ARTICLE INFO

Article history:
Received 20 March 2015
Accepted 31 December 2016

Keywords:
Public sector
Outputs and outcomes
Factorially determined multi-output production
Overall preference effectiveness
Output mix efficiency

ABSTRACT

Key concepts in efficiency analyses are efficiency and effectiveness. Efficiency is popularly connected to 'doing the things right' and effectiveness to 'doing the right things'. The paper elaborates upon the latter concept within a setting where resources are transformed into outputs under the control of a public provider, while outcomes with outputs as inputs represent higher social goals, but this production is outside the public provider's direct control. A new measure of overall preference effectiveness is introduced and its decomposition into output-oriented efficiency and output-mix efficiency is shown. The monumental task of getting the necessary information for calculating effectiveness is highlighted.

1. Introduction

The concepts efficiency and effectiveness are used in many studies, but the bulk of empirical studies provide measures of efficiency, while effectiveness measures are presented much more seldom, and different definitions are used. The standard definitions of efficiency and effectiveness have been expressed by many authors by stating that efficiency is a question of doing things right, and effectiveness is a question of doing the right things (see, among several publications, Drucker [18] in the popular management literature and Fitz-Gibbon and Tymms [24] being typical for journal papers). This definition obviously builds on the assumption that when employing resources to produce 'things' there must be a way of evaluating what are the right 'things' to produce. For production units in competitive industries with prices both on inputs and outputs evaluation follows from the objective function such as profit maximisation or cost minimisation. Considering cost minimisation as a benchmark Farrell [23] showed that his overall efficiency measure, or cost efficiency measure, can be decomposed multiplicatively into relative loss due to 'technical' inefficiency, i.e. not realising the potential at the frontier production function, and relative loss from using a non-optimal mix of inputs termed allocative inefficiency. As to the effect of observed output mix, loss of a non-optimal mix of outputs can be measured analogously by using revenue-maximisation (for given inputs) as the objective function. A unit realising overall efficiency may therefore both doing things right and doing the right things, i.e. fulfilling effectiveness.

However, for service production within the public sector that is not priced and sold on markets the situation is not so straightforward. There is a problem to know or find out what are the right things to do.

The purpose of the paper is to define effectiveness in such a way that a clear distinction between doing things right and doing the right things is made in a formal way. A model for public service production is introduced that distinguishes between efficiency and effectiveness by assuming that the public sector producer controls doing things right only, and that there is a transformation of these things to another type of things that are outside the direct control of the service provider. Things produced by the service provider and things outside the control should then be distinguished by different names, and a natural choice is to use outputs for services directly produced by the provider, while the things outside the control of the producer but evaluated by consumers are termed outcomes. This terminology can be found in many papers (see e.g. Burkhead and Hennigan [8], Bruijn [7], and Schreyer [49]).[1]

A typical situation for the type of public services that I have in mind for the efficiency – effectiveness model being developed is that the service provider is set up to serve more general social

☆The paper is based on parts of Førsund [27] and is written within the research programme "Kostnadseffektiv drift av Forsvaret (KOSTER III)" (Cost Efficiency in Defence) at the Norwegian Defence Research Establishment (FFI). I have benefitted from comments by Espen Berg-Knutsen, Kjell Arne Brekke, Erwin Diewert, Rolf Färe, Torbjørn Hanson, Sverre Kittelsen, Mika Kortelainen, Sverre Kvalvik, Wallace Oates, Jørn Rattsø, and Kenneth Løvold Rødseth. Perceptive and constructive comments by an Associate Editor and two anonymous referees improving the paper are gratefully acknowledged.

☆☆This manuscript was processed by Associate Editor Portela.

* Correspondence address: Department of Economics, University of Oslo, Norway.
E-mail addresses: finn.forsund@econ.uio.no, f.r.forsund@econ.uio.no

[1] The distinction between outputs and outcomes may originate in the health-economics literature [49] and political science and public administration literature [48]. Note that if input and outputs of private providers are traded on competitive markets there is no real distinction between outputs and outcomes. However, if outputs have external effects (negative or positive) not reflected in output prices then social effectiveness still needs a prioritising between outputs.

http://dx.doi.org/10.1016/j.omega.2016.12.007
0305-0483/

objectives than the actual services themselves reflect. Hospitals treat patients of various categories as individuals, but hospitals will also improve the general health of the public at large. Educational institutions provide education of various types serving a higher goal of contributing to the human capital formation. Labour offices provide training courses in various skills and do job searches for unemployed in order to reduce the rate of unemployment as the final goal. Branches of defence like army, air force and navy produce services to serve higher goals like preserving the peace and guarding the independence of a country. Such higher goals may be the reasons for setting up the public service-producing units in the first place, and the goals are usually expressed in statements of the intent of providing services together with the concrete goals of the service-producing units. The societal value of providing services may be expressed by the success of obtaining the higher goals, or the improvement in indices measuring such goals. Higher social goals may typically be associated with outcomes of a public good nature, meaning in a strict sense that realised states of goals once established can be enjoyed by anyone, and one person's consumption does not reduce consumption by other persons. However, the analysis may also be relevant for goods that are useful generally for the public. The assumption that the service provider does not explicitly control the outcomes, and that outputs are provider-specific, but not outcomes, seem to fit well with the assumption of a public good nature of outcomes.

A standard dictionary definition of effectiveness is the one given in Cooper and Ijiri [14]: "Ability to (a) state and (b) achieve objectives." This is also stated in the efficiency literature [9]. Objectives may be profit maximisation, revenue maximisation or cost minimisation, as mentioned for the private sector operating in competitive markets for both inputs and outputs above. It is nothing wrong with this definition, but the emphasis on the type of public service I am studying is that effectiveness involves making explicit prioritising of the services to be produced by a public provider in order to maximise the societal value of the ultimate impact on social objectives.

A contribution of the paper is thus to explicitly show the connection between efficiency and effectiveness within the type of model being developed distinguishing between provider-specific outputs and public good outcomes.

Ultimate goals may be lofty. In order to be operational the outcomes must in principle be measureable and be represented by indicators (see Hatry [35] for examples). The distinction between service outputs and outcomes may be fuzzy. In practical politics the indicators used to measure outcomes may degenerate to indicators of service outputs and vice versa. The ultimate goal for higher education may be an improvement in human capital, but outcomes may also be conceived of as the number of candidates with different types of education. One ultimate objective of defence may be to keep the peace, but measurable outcomes may be the upkeep of national sovereignty, national crisis management, participation in international UN peace force operations, and similar more concrete activities, as stated in official Norwegian document concerning the defence sector [45]. Another problem with goals of the military is that in general either you have peace or not. Thus this outcome can only take two values. In order to value effectiveness of outputs it is therefore necessary to develop indicators for zero – one goals by trying to construct continuous indicators that make the higher goals operational. One approach is to construct scenarios for possible conflict situations and find expressions for how different levels and mix of outputs in terms of military capabilities fulfil the higher objectives (see Hanson [33] for how this can be done for the Norwegian Home Guard).

The plan of the paper is as follows. A brief review of literature distinguishing between outputs and outcomes as well as between efficiency and effectiveness, but done in ways that are in contrast to the approach of the paper, is presented in Section 2. The two types of production relationships for outputs and outcomes of a service provider are elaborated upon in Section 3. The concept of effectiveness of outputs in the provision of outcomes is modelled and discussed in Section 4, and a decomposition of a Farrell-inspired type of overall preference effectiveness measure is presented. Section 5 concludes with emphasis on implications for information requirement for effectiveness analyses.

2. Literature review

An interesting early paper that distinguishes between the services produced and the perception of the services by the consumers is Bradford et al. [6]. There service outputs provided by a public producer are classified as direct outputs ('D-output'), while "the thing or things of primary interest to the citizen-consumer" is termed 'C-output' (p. 186). The C-output is a function of the D-output and environmental variables. An example in Bradford et al. [6] of D-outputs of the police is foot- and car patrols within a district and the C-output being the level of safety felt by inhabitants. Another example there from educational institutions is D-outputs as number of pupil-lessons in various subjects, while C-outputs may be skills obtained measured by e.g. test scores. However, efficiency or effectiveness concepts are not discussed in Bradford et al. [6]. No inefficiency was actually introduced.

Although Bradford et al. [6] have 151 citations in Social Web of Science (per 01.11.2016) few have followed up their distinction between the two types of services in efficiency analyses. Ruggiero [46,47] and Duncombe et al. [19] use explicitly the concepts of D-output and C-output in efficiency analyses of secondary schools. However, the distinction between efficiency and effectiveness is not pursued. It is the role of environmental variables as fixed variables together with discretionary inputs that is explored in a one-stage setting using the non-parametric data envelopment analysis (DEA) to calculate efficiency scores.

As stated in the Introduction the purpose of the paper is to define effectiveness in such a way that the distinction between doing things right and doing the right things is made in a formal way. By extending the distinction in Bradford et al. [6] between D-outputs and C-outputs by letting C-outputs be outcomes of a public good nature that is not controlled directly by the service provider of D-outputs, makes it possible to be explicit about what it means to produce the right things. Some sort of evaluation of outcomes is needed. The implications of the fundamental feature of producing service outputs in order to serve higher social goals for effectiveness measurement can then be explored. To distinguish between service outputs and outcomes turns out to be crucial for how to approach efficiency and effectiveness measurement. I will reserve efficiency for doing things right and use effectiveness to characterise doing the right things in an explicit way using a preference function for evaluating outcomes. The crucial assumption is that the transformation of service outputs into outcomes follows processes that the service output provider does not control directly; the provider can only influence outcomes *indirectly* through its outputs. The relevance of this assumption is, of course, an empirical question, but I will here explore the consequences of making such an assumption for the ability to give clear and distinctive measures of efficiency and effectiveness. A Farrell type of decomposition of cost efficiency into technical efficiency and allocative efficiency will be developed for characterising the mix of outputs in a situation without prices on outputs.

The purpose of the literature review is not to make a standard general review of a field, but to use a limited sample of papers using the terms outputs, outcomes, efficiency and effectiveness to contrast what is typically done in this literature and what is the

Please cite this article as: Førsund FR. Measuring effectiveness of production in the public sector. Omega (2017), http://dx.doi.org/10.1016/j.omega.2016.12.007

approach of this paper. The papers cover applications to education, health, transport, and libraries among others.

A type of approach in the literature is to distinguish between two stages; first services produced by a provider and then a second type of services produced involving actively consumers of the provider's services. When calculating what is termed effectiveness in production at the second stage the service outputs are used as inputs. The problem with this approach is that there is no prioritising of the service outputs in the second stage, thus effectiveness defined as doing the right things does not follow. The way of calculating the measures for the two stages are formally the same. In my approach there are also two stages, but the second stage is not controlled by the public service provider and efficiency and effectiveness of the two stages are intrinsically involved. The public good aspect of outcomes in the second stage is missing from the reviewed papers below; instead the outcomes in the second stage are usually typical private goods. Furthermore, it is often the case that it would be more logical to merge the two stages and just measure standard efficiency in using the resources to produce the outcomes.

The papers are grouped according to activity studied, and the differences between the approach in the paper and the reviewed papers are highlighted.

In a model of education production of secondary schools Lovell et al. [41] distinguish between two levels; i) conversion of school resources into services, and, ii) services converted by students into intermediate and long-term outcomes. However, although efficiency is used for the activity of converting school resources into service outputs and effectiveness is used when outcomes are outputs, the standard efficiency model is used in both cases.

In the health economics literature a natural distinction is often made between the interventions performed on patients and the health improvement, e.g. measured by the difference between health status before and after interventions. Such health changes at the patient level are often termed outcomes. To find the treatment or interventions that lead to the best post-intervention status is then an efficiency question of picking the optimal treatment, while standard efficiency studies usually focus on efficiency of units performing the same interventions. Converting service outputs, defining the interventions, into client outcomes is then termed effectiveness.

In Schinnar et al. [50], studying mental health partial care programs for the fiscal year 1984/1985, the concepts productivity, efficiency and effectiveness are used. It is stated that "the distinction between these concepts rests with the choice of variables used to represent 'inputs' and 'outputs' and not with the method of measurement" (p. 393). The measurement is based on a best practice frontier serving as multi-dimensional reference surface for all units, and the non-parametric DEA method is applied. Thus, productivity is a measure of what is now termed Farrell technical efficiency using various categories of hours of labour as inputs, and efficiency is Farrell's overall or cost efficiency, both measures using services like treatment staff time and administrative staff time spent on clients as outputs, while when measuring effectiveness service outputs are used as inputs and client outcomes as outputs.

Amado and Dyson [1] use data (for a single month) on primary diabetes care to study equity, technical efficiency, cost efficiency, clinical- and patient-focused effectiveness, and equity. A DEA model is used. As in Schinnar et al. [50] cost efficiency is measured for the transformation of financial resources into service outputs like number of patients given different interventions, while service effectiveness is measured by using as inputs the service outputs and using outcomes based on patient satisfaction and number of patients for whom diabetes problems did come under control as outputs.

In Färe et al. [21] a framework for assessing the efficiency of health care, based on DEA models, is introduced distinguishing between the use of standard inputs to produce medical interventions as service outputs and then the outcomes of the interventions as a function of the interventions. The outcomes are defined as the health status of a patient after the interventions and connected to Sen's idea about capabilities to enjoy commodities, i.e. the ability to enjoy the health outcomes, thereby characterising various aspects of better quality of life by an index. However, the distinction between efficiency and effectiveness is not pursued.

The approach of my paper is focused on the problem of prioritising between service outputs provided by an agency when the higher goals, or indices constructed to cover such goals, can only be influenced by the public service provider through the choice of the level and mix of its service outputs for given resources.

Solà and Prior [52] distinguish between efficacy and effectiveness in a study of Catalan hospitals, using dictionary definitions of the two terms. Thus, the former term is defined as achievements of targets, while the latter term is defined "as the degree at which production reaches the final targets" (p. 220). However, only efficiency and productivity measures based on standard DEA efficiency measures are actually computed.

The basic idea in Bradford et al. [6] of distinguishing between the outputs of a service provider and the outputs that consumers enjoy is also found within transportation economics [12,13,53–55], and in studies of efficiency of public libraries [15,16,31], the latter based on insights from administrative science. However, the distinction is between potential service provision (outputs) and the actual services enjoyed by consumers (outcomes). Thus transportation outcomes are defined as the actual use of transportation capacities - bus, metro, railway, and aeroplane – measured by passenger miles or number of passengers transported and ton-miles of freight. Outputs are the *potentials* for providing these services, i.e. number of transport units per time unit and seat miles and ton-miles. The objective of transport activities is to transport people and freight, but the transport companies cannot determine the volume (other than the upper limit set by the capacities).

Analogously, in the case of public libraries the service output is the *potential* for lending out books, while the outcome is the actual lending. Neither in the transportation references nor in the library references is there any reference to Bradford et al. [6], and the library papers have no references to the transportation literature, but in De Witte and Geys [16] there is a list of type of service production with short definitions of service potentials and service delivered for service providers ranging from water utilities to public transport (Table 5, p. 601).

An implication of measuring outcomes this way is that the production of outcomes is uniquely tied to the agency providing the outputs. This may also be the case if the outcome of a teaching institution is measured by the quality of the education measured by the average score of the graduates from that institution, and if outcome of treatments at a hospital is measured by the number of patients that is cured, or the health improvements of its patients.

However, the way the outcome production will be modelled in this paper is more general than found in these strands of literature. A problem with the transportation literature is that service effectiveness is defined and calculated as the Farrell technical efficiency measure using outputs as standard inputs and outcomes as the outputs. A similar exercise is done in the library literature. But our purpose here in measuring effectiveness is to characterise the choice of the output mix under the service provider's control. This is not done by comparing potential services and actual services. It is also the case that investing in capital and setting up capacities are only done based on calculations of demand. It is standard in production theory to regard capacities as capital inputs in a model using realised demand as outputs. Variable factors like labour are not employed for potential production, but for actual production.

Please cite this article as: Førsund FR. Measuring effectiveness of production in the public sector. Omega (2017), http://dx.doi.org/10.1016/j.omega.2016.12.007

In Medina-Borja and Triantis [43] and Medina-Borja et al. [44] a large-scale theoretical and empirical project of evaluating the performance of not-for-profit human and social welfare service organisations is undertaken. A distinction is made between service outputs and customer outcomes ([6] is referred to in [43]). Effectiveness is used in the text to characterise outcome achievements. However, the efficiency for outcome achievement (Table 3, [43]) is calculated by using a DEA model with service delivery as inputs and outcome achievements as outputs, e.g., again there is no explicit question about prioritising between service outputs.

If outcomes are pure public goods it may be the case that the public does not demand the service outputs provided by the agency, but demand the outcomes themselves. An example is the military. The public has preferences for the final outcomes, like keeping the peace, but does not demand troops, exercises, equipment, or the various activities at the service output level. However, there may also be types of public service outcomes that have individual demand, and then our formulation will coincide with C-outputs as defined in Bradford et al. [6]. Individuals demanding service outputs may then be transforming these services into individual outcomes based on a household production function approach a la Becker [3] or Lancaster [39]. However, I am thinking more in terms of outcomes as public goods. Estimation cannot then be based on observations of actions of individual consumers.

A typical feature of the relationship between service outputs and outcomes is that this transformation process is not controllable within a specific production activity. The way from output to outcome is a process happening to individuals consuming (or being exposed to) the service and actions of individuals outside the direct control of the service provider to influence the final outcome.

The concepts of efficiency and effectiveness are used somewhat differently in the DEA literature. The concept of outcome may correspond to objectives in the Cooper and Ijiri [14] definition. However, in the DEA literature objectives have been stated as achieving target levels of outputs and effectiveness used for measuring distance between outputs and target for outputs [28], and using effectiveness when imposing weights on outputs [29]. The latter approach is also followed in Asmild et al. [2] measuring what is there called effectiveness using more general weight restrictions.

In a series of papers [30,37,38] an alternative to restricting weights in DEA was introduced leading to what was termed 'value efficiency' analysis. A decision-maker for a production unit was assumed to have preferences for a specific combination of outputs and inputs on the efficiency frontier estimated by DEA. The inefficient units were measured relative to this most preferred solution. However, the value function was not known, therefore a tangent cone to the point was used as reference. The approach has its roots within multi-criteria decision-making. However, standard objectives in production theory within economics of maximizing profit, minimising costs for given outputs, maximising revenue for given inputs, etc., are completely neglected, although in one application [37] relevant data are available. Our use of a preference function in Section 4 is quite different.

A crucial difference between the approach of the paper and the approach in studies reviewed above within education, health, transport and libraries, is the assumption in the present paper that the public service provider can influence the transformation of resources into service outputs, but cannot directly influence the transformation between service outputs and the outcomes. Furthermore, I go further than the seminal contribution of Bradford et al. [6] in assuming a basic public good aspect of outcomes as a generalisation of "things of primary interest to the citizen-consumer". It follows from these assumptions that estimation of a frontier function to serve as a benchmark for effectiveness using service outputs as inputs and outcomes as outputs is not so relevant. It is the resource use, and the allocation of resources on the different service outputs, that can be influenced by the production unit in question and not directly the outcomes in the present approach. If outcomes can be influenced directly by the service provider then service outputs can be regarded as intermediate outputs. But it is then more relevant (and commonly done) to skip the intermediates and use outcomes as outputs and then calculate efficiency with outcomes as outputs and resources as inputs. In Lovell et al. [41] it is stated that one type of outcome in the educational sector is of a long-term nature and earnings are used as an outcome. However, it seems awkward to assume that a former student's earning can be influenced actively by the educational institution of the student after the student has left.[2] It seems that the outcome of a long-term nature that is a function of the state of the macro economy, etc., is better represented as having a public good nature.

In the health efficiency literature outcome is naturally used for the health status after interventions. But these outcomes are connected to specific patients and are clearly of a private good nature. It may be the case that there is only one way of treating patients, e.g. a specific procedure is followed for a hip replacement operation chosen as the current best practice, and then the outcome can be accurately measured after the recovery of the operation of the patient. In other cases it may be more difficult to both make a diagnosis and to choose treatment. Then outcome indicators may differ between the health care units in question. However, the relevance of first studying the efficiency of the 'internal' procedures with the latter as outputs and resources as inputs, and then studying what is then termed effectiveness using service outputs as inputs and outcomes as outputs can be questioned; after all it is a question of utilising the resources to produce the best outcomes. The inefficiency stems from the transformation of resources into interventions as service outputs.

In the transport and library sectors there is another type of awkward use of the effectiveness concept. This seems to be tied to how a unit utilises its capacity to produce services that are actually consumed by individuals. But there may be rational reasons for operating with capacities that have a different rate of utilisation. It depends e.g. on the policy of how to serve peak load periods of job-related commuting morning-afternoons, and the quality level chosen, varying from seats for everyone (as is the practice in some long-distance train and bus services) to being packed standing as sardines in a can.

In the literature reviewed above the efficiency problem and the effectiveness problem are set up as two separate operations. However, the approach used in the paper leads to the efficiency and the effectiveness problems to be highly interconnected.

A recent paper by Mayston [42] studies a public sector university that can leverage up the exogenously given base funding by increasing quality and volume of outputs by attracting better faculty generating more external funds due to higher quality of research, and attracting better students by better teaching quality making increasing the fees possible. Thus the budget is endogenised. Effectiveness is connected to the additional performance made possible by these endogenous effects. Efficiency is used to characterise the production function in the variables inputs, volume of outputs, the quality of outputs, and socio-economic variables impacting the productivity of inputs in the standard way. Assuming that the available budget is exhausted the impacts on outputs and quality of extending the resource base can be traced. Some interesting new concepts are introduced to characterise the model set-up and possible results. However, the focus on revenue increases induced through quality changes are quite different from

[2] It is also a problem with the time period of the education and of the earnings; obviously the earning periods come after the education period and there is no active interaction student-institution any more. The influence is through the initial quality of the education or reputation of the institution.

Please cite this article as: Førsund FR. Measuring effectiveness of production in the public sector. Omega (2017), http://dx.doi.org/10.1016/j.omega.2016.12.007

my approach of focussing on the public good aspect of public service providers. The fundamental distinction between outputs and outcomes is not there, and the term effectiveness is not used for characterising doing the right things within a given budget constraint. However, some results like the decomposition of the overall effectiveness terms in Mayston [42] have a formal likeness with the decompositions in the present paper.

3. The production relationships of an agency

Let me call a service-producing unit in the public sector an agency. A single agency only is considered (suggestions for a generalisation are indicated in Eq. (3) below). The multi-output nature of service production can be modelled in several ways, from very general formulations of a transformation function in multiple outputs and multiple inputs to more specialised formulations taking care of technical connections between outputs. Such formulations may involve independent and parallel activities for each output in chains of intermediate deliveries ending in the final service delivered to the consumers, or resources may be shifted around to produce any of the services. Concerning service production in the public sector the main input, at least in terms of current costs, will typically be labour of different qualities. Real capital may represent substantial investments in specialised buildings and machines like in hospitals, but in many cases capital is generic, like office buildings and computers.

It seems reasonable to model a great deal of flexibility as regards the possibility of what mix of services to produce given the inputs. Therefore a standard transformation relation between outputs y and inputs x (y and x are interpreted as vectors) seems appropriate, assuming that relations between variables can be found using the implicit function theorem:

$$F(y_1,\ldots,y_K,x_1,\ldots,x_N;z_F)=0,\partial F/\partial y_k\geq 0,\partial F/\partial x_n\leq 0,\quad k=1,\ldots,K,n=1,\ldots,N \tag{1}$$

There are N types of resources $(x_1,\ldots,x_N)$ and K types of services or outputs $(y_1,\ldots,y_K)$ produced by an agency. The signing of the derivatives follows standard convention, as is also the case of normalising the right-hand-side constant to zero. Inputs can freely be allocated to any mix of outputs, implying a maximal *degree of assortment* of outputs [26]. The vector z_F represents variables that influence the relationship between inputs and outputs, but these variables are non-discretionary and will in our analysis be regarded as exogenous (uncontrollable) for an agency (symbolised by using a semi-colon in front of the vector). Such variables, also termed environmental variables in the literature, may occur for the type of services where the ultimate consumers are present in the production process like students in higher education and patients in hospitals, as mentioned above. Socio-economic background and inherent capabilities of students may be examples of exogenous variables in higher education.

In order to model how service outputs influences the objectives measured by the outcomes another type of production relations than (1) may be informative. The outcomes cannot be controlled directly by the agency; it can only observe (in principle) the outcomes influenced by its production of services (given the values of the exogenous variables). It then seems appropriate to use the special multi-output relationships of a type Frisch [26] termed *factorially determined multioutput production*. There are M final outcomes Y that each are functions of the K agency service outputs y:

$$Y_m=g_m(y_1,\ldots,y_K;z_{g_m}),\frac{\partial g_m}{\partial y_k}\geq 0,m=1,\ldots,M,\quad k=1,\ldots,K \tag{2}$$

The instruction to the agency from the financing institution (e.g. a parliament, government or ministry) concerns the service outputs y. The financing institution is assumed to have some knowledge about the impact of service outputs on outcomes Y when formulating goals in terms of service outputs, but typically not a complete understanding. Therefore the agency only concerns itself with the provision of outputs y and does not act on any information about outcomes Y, as different from the two stage approaches in the reviewed literature in the previous section.

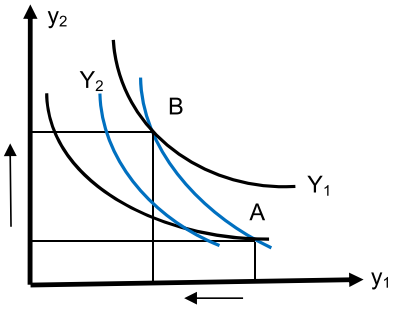

Fig. 1. Factorially determined multioutput production of outcomes Y with outputs y as inputs.

The outcomes are separable in the sense that each outcome can be expressed as an outcome-specific function with the same set of service outputs as arguments. (The z_{g_m} variables will be commented upon below.) In Frisch [26] this is termed *product separation*. It is a special kind of separation in the sense that the degree of assortment is zero, meaning that for given outputs all the M outcomes are determined. However, when varying the outputs different proportions of outcomes may be realised. Each outcome has its unique set of isoquants in the common input (i.e. the agency output) space illustrated in Fig. 1 in the case of two outcomes as products and two service outputs as inputs. The two sets of isoquants for the two products are drawn in the common input space. Moving along the isoquant for Y_2 from point A to B, reallocating inputs in order to increasing the amount of output y_2, here functioning as an input, and reducing the amount of output y_1 as an input, the outcome Y_1 is increased and Y_2 kept constant.

It should be stressed that introducing a production function like (2) at this stage is more a conceptual and abstract idea than a description of production relationships that can be considered as well-defined in a technical sense (cf. the statement in [8] that "in the public sector there is almost no production function that can be conceptualized with clarity").

In addition to the controllable service outputs y we have also opened the possibility for other environmental (or uncontrollable) variables z_{g_m} (interpreted as a vector) influencing the outcomes. The general level of health in the population does not only depend on service output from hospitals, but also on individual characteristics such as smoking, obesity and other lifestyle variables. The formation of human capital does not only depend on the number of exams taken, but also on the quality of students concerning development after leaving educational institutions. Notice that the exogenous variables z_F and z_{g_m} in the two production functions are not necessarily the same.[3] If the general objective is to reduce the occurrence of criminality in the population other factors beside the services provided by the police will influence this. The defence objectives of keeping peace and independence of a country are highly influenced by actions of other countries. These environmental (or uncontrollable) variables are assumed constant throughout the paper.

It is a question how to handle quality in the current context. In principle the service outputs are measureable, although the proper treatment of quality may be difficult to capture as quantitative

[3] In [47] it is stated that exogenous or environmental variables enter both stages referring to D-outputs and C-outputs introduced in Bradford et al. [6].

Please cite this article as: Førsund FR. Measuring effectiveness of production in the public sector. Omega (2017), http://dx.doi.org/10.1016/j.omega.2016.12.007

measures. A standard procedure is then to assume the same unknown quality level for all providers of the same type of service. This treatment of quality is, of course, not satisfactory. One approach may be to assume a multiplicative decomposition of the service into a quantity part, i.e., number of tax returns processed, and a quality part catching the accuracy of the work. It is reasonable to assume that better quality requires more resources [11,52], e.g., for the same number of tax returns processed more labour has to be used. There is a trade-off between numbers of tax returns processed and the quality of the work. The level of quality to go for then has to be determined by the service provider. Although this line of reasoning is very interesting to develop, it will not be considered in detail in the paper.[4] Quality aspects of outputs may be entered in the output vector y controlled by the agency [42].[5] The outcomes are then dependent on both output quantity and output quality. It is also possible to consider quality variables of the outcomes Y that are different from the output quality variables.

Considering several agencies n producing the same type of outputs the arguments in the outcome production functions may either be the sum of the outputs from all agencies providing the same outputs, or just the outputs provided by the specific agency. In the former case we have

$$Y_m = g_m\left(\sum_{j=1}^{n} y_{1j}, \ldots, \sum_{j=1}^{n} y_{Kj}; z_{g_m}\right), \quad m = 1, \ldots, M \tag{3}$$

This will be the specification in the case of the outcomes having the character of a public good, like the outcomes for defence. However, since the marginal impact of outputs on outcomes is the same for the same type of output independent of agency I will only operate with a single agency for convenience.

It may be the case that the service outputs of several agencies being different influence the same objectives measured by outcomes Y. It may also be the case that services from one agency having a positive effect on the agency's own objectives have negative effects on objectives of other agencies. One example of conflict may be efficiency-related outcomes and outcomes based on distributional objectives. Another example of a negative effect of an output may be the impact on keeping the peace for a country of participating in military actions in other countries regarded as an output. The latter activity may create reactions involving terrorist attacks (blow-back), as we have seen happening in USA, England and recently in France and Belgium.

4. Effectiveness in the provision of outcomes

The two types of variables outputs and outcomes are not dated in (1) and (2) implying that they occur simultaneously. The agency cannot control directly how resource use influences outcomes, only indirectly through the outputs provided. However, from a social point of view we are interested in efficient use of resources; after all the resources have alternative uses. Therefore we are interested in efficiency of the two different activities; efficiency in the resource use of producing service outputs, and in achieving effectiveness by choice of mix of outputs. These two efficiency considerations are somewhat different. We assume that the N resources have well-defined prices q_n $(n=1,\ldots, N)$; inputs are bought in competitive markets. But typically service outputs are not sold in markets, and concerning outcomes they are more or less by definition not priced in any market. Therefore the question whether the right type of service is produced for the relevant outcome may need another type of approach than when studying efficiency in producing service outputs only.

The relationship $F(y, x; z_F) = 0$ is used as a yardstick or benchmark for the most efficient way to combine inputs to produce given levels of outputs. An agency is efficient if $F(y, x; z_F) = 0$, and inefficient if $F(y, x; z_F) < 0$ [32], where y and x are the observed vectors of outputs and inputs of an agency and z_F a vector of exogenous (uncontrollable) environmental variables influencing the transformation of resources into service outputs. In the efficiency literature $F(y, x; z_F) = 0$ implicitly defines the frontier function and efficiency measures are calculated for observations relative to this frontier by solving for the minimal scalar E_2 in $F(E_2^{-1}y, x; z_F) = 0$, where the common adjustment factor E_2^{-1} is the inverse of the Farrell [23] measure of output-oriented efficiency E_2. The popular data envelopment analysis (DEA) estimates a non-parametric piecewise linear frontier and yields the efficiency scores [10,23,25].

When output prices do not reflect consumers' evaluation the introduction of a preference function is necessary in principle in order to be able to prioritise between the outcomes, and thereby enabling a prioritising between outputs (cf. [8] stating: "The ultimate objective function – that which is to be maximized – should be described as a social state: are citizens better or worse off as a result of a particular government service delivery?"). I have assumed that the measurable outcomes Y are related to the ultimate objectives of providing public services. A social preference function $W(Y_1, \ldots, Y_M)$ $(W'_{Y_m} > 0 \ \forall m)$ based on the links between the ultimate objectives and the measurable outcomes can then be introduced. We are looking at a single agency.[6] If a number of agencies producing the same outputs are considered the choice between the specifications (2) and (3) will influence the modelling.

As is the procedure in economics a social planner is introduced that has a preference (welfare) function over the outcomes and know how inputs are transformed to outputs and outputs transformed into outcomes.[7] Returning to the output quality issue raised in Section 3 the vector y may contain quality variables that will influence the outcomes. The specification of the outcomes Y may also be done separating volume and quality, and outcome quality will then be evaluated in the $W(.)$ function. The values of the environmental z-variables are also assumed known. The planner's job is then to derive the optimal conditions for the priority of outputs by assuming a given budget, B, for the resources x at prices q. It will be assumed that the given budget is the same as the observed costs (i.e. no waste or savings of budget). The optimisation problem is:

$$\begin{aligned} &Max \quad W(Y_1, .., Y_M) \\ &\text{subject to} \\ &\sum_{n=1}^{N} q_n x_n \le B \\ &Y_m = g_m(y; z_{g_m}), \quad m = 1, \ldots, M \\ &F(y, x; z_F) \le 0 \end{aligned} \tag{4}$$

The variables y, x, z_{g_m} and z_F are interpreted as vectors. It seems reasonable to enter the relations between outcomes and outputs using equalities, because the production relations are autonomous in the sense that the transformation process is not under the control of any agency. For services consumed by individuals the

[4] In Färe et al. [22] input and output qualities for secondary schools are introduced as independent variables in a study of productivity.

[5] In Mayston [42] quality is specified explicitly as a separate term analogous to the volume of output within a Cobb-Douglas type of transformation function with fixed partial elasticities.

[6] In order to estimate a frontier function several agencies are necessary. However, a benchmark frontier is usually assumed to be common for all units.

[7] Mayston [42] introduces an evaluation function in the outputs and qualities.

Please cite this article as: Førsund FR. Measuring effectiveness of production in the public sector. Omega (2017), http://dx.doi.org/10.1016/j.omega.2016.12.007

transformation of service outputs to outcomes takes place within the consumers themselves (c.f. household production functions) and can be expressed as an aggregate for relevant groups of consumers and can then be regarded as public goods. Exogenous variables of type z_{g_m} and z_F influencing the process can also act at an individual level, e.g., the state of health of a person treated by the health system may depend on whether the person smokes, and also other lifestyle factors, including exposure to air pollution. So effectiveness is not connected to the production relations $g_m(.)$ not being realised, cf. the equality in (4). However, as shown by the inequality in the case of transforming inputs into outputs it is opened up for the possibility that the benchmark frontier function $F(y, x; z_F)$ may not be realised by an agency.

The Lagrangian function for problem (4), inserting the outcome production functions into the preference function for simplification, is

$$\begin{aligned} L = {} & W(g_1(y;z_{g_1}),\ldots,g_M(y;z_{g_M})) \\ & -\beta\left(\sum_{n=1}^{N} q_n x_n - B\right) \\ & -\gamma F(y,x;z_F) \end{aligned} \tag{5}$$

The necessary first-order conditions are

$$\begin{aligned} & \frac{\partial L}{\partial x_n} = -\beta q_n - \gamma F'_{x_n}(y,x;z_F) \le 0\ (=0 \text{ for } x_n > 0), \quad n = 1,\ldots,N \\ & \frac{\partial L}{\partial y_k} = \sum_{m=1}^{M} W'_m g'_{mk}(y;z_{g_m}) - \gamma F'_{y_k}(y,x;z_F) \le 0\ (=0 \text{ for } \quad y_k > 0)\,, \\ & \quad k = 1,\ldots,K \\ & \beta \ge 0\ \left(=0 \text{ for } \sum_{n=1}^{N} q_n x_n < B\right) \\ & \gamma \ge 0\ (=0 \text{ for } F(y,x;z_F) < 0) \end{aligned} \tag{6}$$

These are the conditions for the social optimal solution. Another question is what an agency is actually doing. The point is that the social solution is the reference for judging the efficiency and effectiveness of an agency.

Let us first look at the possibility of corner solutions for the decision variables resources and agency outputs. The shadow price β translate from the money unit of the budget B to the units of the preference function, showing the increase in the preference-function value of a marginal increase in the budget. The shadow price γ shows the increase in the value of the preference function of a marginal expansion of the production possibilities. By the nature of our optimisation problem (i.e. assumptions about preference and transformation functions), in order to realise the maximal value of the preference function a full utilisation of both the budget and being on the production frontier is necessary; from the complementary slackness conditions for the shadow prices we will typically have both shadow prices positive. The first condition in (6) tells us that a resource should not be used if the value of its partial derivative of the transformation function cannot meet the cost of the factor (value and cost are measured in preference-function units per input unit). Due to the maximal degree of assortment [26] implied by the transformation function $F(.)$ we will typically have that all resources will be fully utilised. For all of the K outputs to be produced we must have that the sum of the preference-weighted marginal productivities of an output is equal to the valuation of the partial derivative of the transformation function with respect to the output in question for each type of output. It is difficult to see a situation where the value of the derivative of the transformation function evaluated using the common shadow price on the transformation-function constraint is greater than the actual value created by marginally increasing the production of the output in question. However, the functional forms of the system of Eq. (2) may create problems for the existence of a unique solution.

Assuming interior solutions for inputs and outputs, eliminating the Lagrangian parameter γ for the transformation function yields:

$$\frac{1}{\beta}\sum_{m=1}^{M} W'_m g'_{mk}(y;z_{g_m})\frac{-F'_{x_n}}{F'_{y_k}} = q_n, k = 1,\ldots,K, \quad n = 1,\ldots,N \tag{7}$$

The condition (7) tells us that an optimal use of a resource x_n is characterised by the cost of a unit of the resource being equal to the benefit it creates in terms of an evaluation of the final outcomes Y through the production of a service y_k. The second term on the left-hand side is the marginal productivity of resource x_n in producing service y_k. The first term is the evaluation of the outcomes generated at the margin by the service y_k. The measuring unit on the left-hand side is money per unit of resource n (the measuring unit of $1/\beta$ is money per preference-function unit). Using the Frisch system of factorially determined multi-outcome production we have to sum over all the outcomes that are influenced by the marginal change in the service y_k. The condition tells us that the monetised value created by employing a unit of a resource x_n to produce service y_k is equal to the unit resource price.

Eliminating both Lagrangian parameters yields the conditions for rates of transformation and substitution for inputs and outputs:

$$\frac{F'_{x_n}(y,x;z_F)}{F'_{x_r}(y,x;z_F)} = \frac{q_n}{q_r}, \quad n,r = 1,\ldots,N \tag{8}$$

$$\frac{\sum_{m=1}^{M} W'_m g'_{mk}(y;z_{g_m})}{\sum_{m=1}^{M} W'_m g'_{ml}(y;z_{g_m})} = \frac{F'_{y_k}(y,x;z_F)}{F'_{y_l}(y,x;z_F)}, \quad k,l = 1,\ldots,K \tag{9}$$

Eq. (8) shows that the optimality conditions in (6) implies allocative efficiency of the inputs, i.e., the marginal rate of substitution between inputs n and r should be set equal to the factor price ratio.

Considering a pair of outputs, y_k and y_l, combining the second condition in (6) for each output, the condition for prioritising right is expressed by Eq. (9). The preferences are over the outcomes Y. To clarify the implications of prioritising right for outputs y we need to see the implications of preferences for outcomes generated by the outputs. Considering changes in two outputs y_k and y_l total differentiation of the preference function yields:

$$\begin{aligned} & \sum_{m=1}^{M} W'_m g'_{mk}(y;z_{g_m})dy_k + \sum_{m=1}^{M} W'_m g'_{ml}(y;z_{g_m})dy_l = 0 \Rightarrow \\ & \frac{dy_l}{dy_k} = -\frac{\sum_{m=1}^{M} W'_m g'_{mk}(y;z_{g_m})}{\sum_{m=1}^{M} W'_m g'_{ml}(y;z_{g_m})}, \quad l,k = 1,\ldots,K \end{aligned} \tag{10}$$

We will call this ratio the *marginal preference rate of substitution* between outputs y_k and y_l. It combines the preferences for outcomes with the properties of the outputs as arguments in the system of production functions (2) for outcomes. The marginal productivity of an output is weighted with the marginal preference impact for each of the outcomes affected. The measuring unit for the total expression is then in preference-function units per unit of output. The values are conditional on the values of the exogenous variables z_{g_m} and the budget B. The condition for priority or mix efficiency is that the marginal rate of transformation between two outputs y_k and y_l is equal to the ratio of marginal preference rate of substitution.

Please cite this article as: Førsund FR. Measuring effectiveness of production in the public sector. Omega (2017), http://dx.doi.org/10.1016/j.omega.2016.12.007

4.1. Effectiveness measure

It is clear from (6) that a point on the frontier function for resources and outputs must be realised for a maximum of the preference function to be obtained. In order to develop an effectiveness measure I will introduce an observation (y, Y, x; z) that have the same budget B as in (4). Combining the observation with an optimal solution (y^*,Y^*, x^*;z) of problem (4) measures can be formulated for potential welfare improvement by forming efficiency measures and decomposition inspired by the Farrell [23] efficiency measures. The overall efficiency for a given budget B may be termed *Overall Preference Effectiveness*, *OPE*, and can be multiplicatively decomposed into two components:

$$\underbrace{OPE=\frac{W(Y_1,\ldots,Y_M|B)}{W(Y_1^*,\ldots,Y_M^*|B)}}_{\text{Overall preference effectiveness}} = \underbrace{\frac{W(g(y;z_g)|x)}{W(g(y^F;z_g)|x)}}_{\text{Preference based output−oriented efficiency}} \times \underbrace{\frac{W(g(y^F;z_g)|x)}{W(g(y^*;z_g)|x^*)}}_{\text{Preference based output mix efficiency}} \tag{11}$$

The first term on the right-hand side of the equality is conditional on the observed resources x, both in the numerator and the denominator, while the second term is conditional on the observed resources in the numerator, but on the optimal mix of resources x^* in the denominator. The improvement of the value of the preference function comes from two sources: realising the frontier transformation function for outputs and inputs by eliminating output inefficiency by proportionally increasing the outputs to $y^F=E_2^{-1}y$, where $E_2\in(0,1]$ is the Farrell output-oriented technical efficiency measure satisfying $F\left(E_2^{-1}y,x;z_F\right)=0$ (see second paragraph of this section), and by changing the output mix of the proportional frontier projection y^F to the optimal mix of y^* on the frontier by reallocating inputs x to x^* in the $F(.)$ function, where $g(.)$ is the vector of M functions $g_m(.)$. It seems reasonable to use the observed inputs, indicated by the notation x when the output point is moved radially to the output production frontier, but using the optimal mix x^* of inputs when changing the output mix to the optimal y^*. Overall preference effectiveness for a given budget B not only assumes that efficiency in producing outputs is obtained, but also that effectiveness is achieved by providing the most potent mix of outputs. The first term on the right-hand side reflects doing things right, and the second term doing the right things.

The observed input vector may be different in composition from the optimal input vector x^*, but the total budget is the same by assumption. For fixed input prices the budget B is a linear aggregation of the inputs to an input bundle. However, the marginal productivities of inputs in the function (1), $(-F'_{x_n}/F'_{y_k})$, will typically be different for the input vectors of the two mixes.

The situation can be illustrated looking at a pair of outputs, y_k and y_l, set out in Fig. 2. We have an observation (y, x; z_F) and a given expenditure B on inputs. This budget is also kept in the optimisation problem (4) so we have $\sum_{n=1}^{N} q_n x_n=\sum_{n=1}^{N} q_n x_n^*=B$. The transformation between outputs y_k and y_l at the frontier is shown by the transformation curve labelled $F(y, x; z_F)=0$ for the initial bundle x of inputs. The curve labelled W going through the point y is not strictly speaking a contour curve of the preference function $W(.)$, but is the curve defined by the marginal preference rate of substitution between outputs y_k and y_l in (8) in the case of two outputs, letting y_k and y_l vary in such a way that the value of the preference function along this curve is fixed at W, i.e. the induced changes in outcomes by varying the outputs must be such that the value of the preference function is fixed. A contour curve for the function $W(.)$ will be in the outcome space of Y, while we

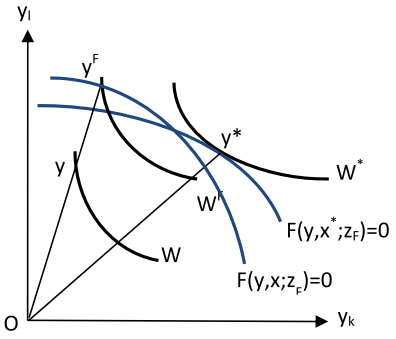

Fig. 2. Effectiveness and its components.

are now operating in the output space of y. To find the shape of a curve in the latter space may be rather complicated, so a stylised case is assumed here. Moving proportionally to the frontier to eliminate inefficiency in the production of outputs for the given inputs x point y^F is realised. But the solution to the optimisation problem (4) for the outputs implies another mix than y^F, namely y^*. The curve labelled W^F, defined the same way as described above, implicitly determined by the properties of the preference function and the outcome production functions, passing through the frontier output point y^F, has a smaller value than the curve W^*, determined from the right-hand side of (7) (by keeping the value of the preference function fixed at W^* but varying y_k and y_l) being tangential to the frontier with x^* as inputs at point y^*. Therefore we move from point y^F to point y^*. This is the realisation of mix efficiency. Effectiveness is achieved by producing the optimal mix of outputs yielding the maximal value of the preference function for a given budget. The production possibility sets will differ for different mix of the inputs for a constant budget. Comparing the vectors x and x^* some inputs will decrease, other increase to keep the budget constant. As stated above the marginal productivities of inputs in the function (1) will be different for the two mixes, as illustrated in Fig. 2 by the two transformation curves labelled $F(y, x; z_F)$ and $F(y, x^*; z_F)$, respectively. In order to understand Fig. 2 it may help to introduce the concept of *cost-indirect output set* [51]. By construction this set will envelope all production possibility sets spanned by input vectors obeying the budget constraint in (4) [20,21], i.e., the coordinates (y^*, x^*) will always be on the cost-indirect frontier.

We notice that both the output efficiency term and the output mix term are derived simultaneously, in contrast to the literature reviewed in Section 2 treating efficiency and effectiveness as two separate problems. However, we see that the output efficiency term uses the Farrell output-oriented efficiency measure E_2 in its calculation. There is a discussion in the literature about the connection between efficiency and effectiveness as calculated using two separated models with empirical results showing both positive and negative correlations (see Kerstens [36], Borger et al. [5], and references to this literature given in these papers). One reason for these results may be the fact that using observed service outputs as inputs in the effectiveness calculation having outcomes as products, the observations on service outputs contains inefficient output levels. The more inefficient a unit is the smaller is the input in these calculations, but, *Ceteris Paribus*, this leads to a higher effectiveness score. The output data are not cleaned up, but remain 'contaminated' with inefficiency as opposed to our approach in (11) where observed outputs are projected to the frontier of the output production function before the efficiency mix score is calculated.

However, although the decomposition of the effectiveness measure is inspired by a Farrell-type decomposition of revenue efficiency (see e.g. [20], pp. 113–115), the terms do not have the

Please cite this article as: Førsund FR. Measuring effectiveness of production in the public sector. Omega (2017), http://dx.doi.org/10.1016/j.omega.2016.12.007

same interpretation in our setting of maximising the value a preference function over outcomes for a given budget. The output mix efficiency term in (11) does not coincide in general with the mix efficiency term that would appear in a revenue- maximising problem because the input mix is not given but free to change under the budget constraint.[8]

The first term on the right-hand side in (11) (after the second equality sign), is in general not identical to the output-oriented Farrell technical efficiency measure E_2 for the same data and frontier function. Output-oriented efficiency in (11) is calculated using values of the preference function. The measures will coincide if both the outcome production functions are homogeneous of degree 1 in the outputs and the preference function is homogeneous of degree 1 in the outcomes:[9]

$$\frac{W(g(y;z_g)|x)}{W(g(y^F;z_g)|x)}=\frac{W(g(y;z_g)|x)}{W(g(E_2^{-1}y;z_g)|x)}=\frac{W(g(y;z_g)|x)}{W(E_2^{-1}g(y;z_g)|x)}=E_2 \tag{12}$$

If this is the case the contour curves illustrated in Fig. 2 will be radial projections of each other, and the spacing of the contour curves is constant in relative terms.

The marginal preference weights in (7) are variable. If it is assumed that the marginal weights are constants, this is equivalent to the preference function $W(.)$ being linearized; $W'_m(Y_1,\ldots,Y_M)=w_m(m=1,\ldots,M)$. Such constant valuation coefficients may play the role of 'prices' of the outcomes. But notice that such prices relate to implicit prices of outputs in a complicated way involving the production functions (1).

The question of how to construct preference functions for public sector outcomes is a research field in itself. There is a literature focusing on how to construct scalar-valued objective functions for macro-economic decision models. Pioneers were the first joint Nobel Prize winners Frisch and Tinbergen (see the account of the ideas of Frisch of establishing preference functions by interviewing decision-makers in [4]).

An example of linearizing a preference function over outcomes is found in Lauer et al. [40] based on works of WHO of performance ranking of health systems of 191 member countries. Five outcome variables for the health sector of a country are used; level of population health, inequalities in health, level of responsiveness, inequalities in responsiveness and fairness in financial contributions. The establishment of fixed weights was based on responses to a survey by over 1000 health experts.

In the literature there are examples of just a single outcome ([6] mention safety level for the police sector and average scores for schools). Then there is no preference function to be maximised, just the index for the single outcome. However, the problem of prioritising between the outputs remains. The contour curve in Fig. 2 will be an isoquant of the single outcome production function of type (2). In that case the *OPE* measure reduces to an outcome effectiveness measure, *OE*:

$$\underbrace{OE=\frac{Y|B}{Y^*|B}}_{\text{Outcome effectiveness}}=\underbrace{\frac{g(y;z_g)|x}{g(y^F;z_g)|x}}_{\text{Output-oriented effficiency}}\times\underbrace{\frac{g(y^F;z_g)|x}{g(y^*;z_g)|x^*}}_{\text{Outputmix efficiency}} \tag{13}$$

where the $g(.)$ function is now the production function for a single outcome.

[8] In the solution to problem (4) the optimal input vector will in general be different from the observed input vector (keeping the budget the same) while in the revenue maximising problem with known output prices the input vector is kept constant, and in (4) values of the outcome preference function is used for evaluation and not output prices that may be used by a service provider selling outputs in a market.

[9] This is a sufficient condition; a referee pointed out that an outcome production function that is homogeneous of any degree $r>0$ when combined with a preference function that can be transformed monotonically into one that is homogeneous of degree $(1/r)>0$ will yield the same result.

Calculating effectiveness is not so simple as calculating cost- or technical efficiency for use of resources to produce outputs. The informational requirement is quite formidable. We must be able to define outcomes in the first place. Then we have to know not only the preference function over outcomes (in the case of more than one outcome), but also how outcomes are influenced by service outputs and other exogenous variables. This last task is quite another exercise than determining the transformation function involving resources and service outputs. Diewert [17], when addressing methods for measuring prices of nonmarket goods, states that the most desirable method is some form of purchaser valuation. A general equilibrium approach for the economy embedding public service outputs is suggested as a way of obtaining user based evaluations. However, he comments that the information required to implement such an approach is "just too great" (p. 181). Thus, the method is declared theoretically sound, but not practical.

5. Conclusions

Both the concepts of efficiency and effectiveness are often used in the literature dealing with efficiency. I have tried to make the distinction between these concepts operational by using the terms outputs and outcomes based on the consideration of the degree of control a public service producer has over its production activity. The apparatus of production theory works best when dealing with resources transformed into service outputs under the control of the organisation in question. Outcomes in this paper represent some higher social goals than outputs and are determined by the outputs and other exogenous variables, but these latter and the outcome production processes will typically be outside the direct control of the organisation.

The relationship between outcomes and outputs and variables not under the control of the service provider, is cast within a framework based on Frisch's scheme of factorially determined multi-output production with outputs and non-discretionary variables as inputs. In order to be able to measure effectiveness in the choice of outputs, i.e., calculate a measure of output mix efficiency; we must have some kind of evaluation of the outcomes. Introducing a preference function over outcomes optimality conditions for providing an effective output mix for a given resource budget are derived. It is shown that the measure for overall preference effectiveness can be multiplicatively decomposed into an output-oriented efficiency of realising a frontier technology for the transformation of resources to outputs, and the output mix efficiency of reallocating the use of resources so the optimal mix of outputs is produced. The decomposition highlights that output efficiency and outcome effectiveness cannot be solved as separate problems as done in the literature, but must be handled simultaneously. Furthermore, the output efficiency term in the decomposition of preference effectiveness is not in general equal to the Farrell output-oriented efficiency measure.

The rather monumental task of providing the necessary information for calculating effectiveness based on doing things right and doing the right things is highlighted. A preference function over outcomes must be established, if the organisation in question produces outputs influencing more than one outcome, and also the production relations between outcomes on one hand and outputs and exogenous variables on the other must be established. An additional complication is to capture the links between service outputs, exogenous variables and outcomes. Quite complicated dynamic relationships involving time lags may have to be modelled and uncertainties about future values of exogenous environmental variables considered. As to estimation of relationships there is the question of introducing stochastic environmental variables and variables causing inefficiencies. Generalising the approach to more than a single agency

Please cite this article as: Førsund FR. Measuring effectiveness of production in the public sector. Omega (2017), http://dx.doi.org/10.1016/j.omega.2016.12.007

as pursued in this paper may not be so straightforward, and possible differences in outputs and effects on outcomes must be dealt with.

The social planner is assumed to have full and perfect knowledge about all relationships and exogenous variables. This is clearly unrealistic, but serves here to establish a frame of reference. How to acquire relevant knowledge and how this information is diffused in a multi-agency world is a challenging research task.

As far as I know the approach of the paper has not been attempted in the literature. However, empirical applications based on the approach of the paper are under way for the Norwegian defence sector at the Norwegian Defence Research Establishment (FFI) [34], and at the Institute for Transport Economics Norwegian Centre for Transport Research (TØI), studying efficiency and effectiveness changes of decentralising decisions on public infrastructure transport investments in Norway.

Given the difficulties met when trying to operationalise doing things right and doing the right things in order to distinguish between efficiency and effectiveness, it is understandable that empirical applications of measuring efficiency within the public sector have been limited to transformation of resources into outputs within a process controlled by the service provider.

References

[1] Amado CAF, Dyson RG. Exploring the use of DEA for formative evaluation in primary diabetes care: an application to compare English practices. Journal of the Operational Research Society 2009;60(11):1469–82.

[2] Asmild M, Paradi JC, Reese DN, Tam F. Measuring overall efficiency and effectiveness using DEA. European Journal of Operational Research 2007;178:305–21.

[3] Becker GS. A theory of the allocation of time. Economic Journal 1965;75:493–517.

[4] Bjerkholt O, Strøm S. Decision models and preferences: the pioneering contributions of Ragnar Frisch. In: Tangian AS, Gruber J, editors. Constructing and applying objective functions. Lecture Notes in Economics and Mathematical Systems 510. Berlin – Heidelberg – New York: Springer-Verlag; 2002. p. 17–36.

[5] Borger de B, Kerstens K, Costa A. Public transit performance: what does one learn from frontier studies? Transport Reviews 2002;22(1):1–38.

[6] Bradford DF, Malt RA, Oates WE. The rising cost of local public services: some evidence and reflections. National Tax Journal 1969;22(2):185–202.

[7] Bruijn H de. Performance measurement in the public sector: strategies to cope with the risks of performance measurement. The International Journal of Public Sector Management 2002;15(7):578–94.

[8] Burkhead J, Hennigan P. Productivity analysis: a search for definitions and order. Public Administration Review 1978;38(1):34–40.

[9] Charnes A, Cooper WW. Preface to topics in data envelopment analysis. Annals of Operations Research 1985;2(1):59–94.

[10] Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision making units. European Journal of Operational Research 1978;2:429–44.

[11] Chilingerian JA, Sherman HD. Managing physician efficiency and effectiveness in providing hospital services. Health Services Management Research 1990;3(1):3–15.

[12] Chiou Y-C, Chen Y-H. Route-based performance evaluation of Taiwanese domestic airlines using data envelopment analysis. Transportation Research Part E 2006;42(2):116–27.

[13] Chu X, Fielding GJ, Lamar BW. Measuring transit performance using data envelopment analysis. Transportation Research Part A 1992;26(3):223–30.

[14] Cooper WW, Ijiri Y. Kohler's dictionary for accountants. sixth edition. Englewood Cliffs, NJ: Prentice-Hall; 1983.

[15] De Witte K, Geys B. Evaluating efficient public good provision: theory and evidence from a generalised conditional efficiency model for public libraries. Journal of Urban Economics 2011;69(3):319–27.

[16] De Witte K, Geys B. Citizen coproduction and efficient public good provision: theory and evidence from local public libraries. European Journal of Operations Research 2013;224(3):592–602.

[17] Diewert WE. Measuring productivity in the public sector: some conceptual problems. Journal of Productivity Analysis 2011;36(2):177–91.

[18] Drucker P. An introductory view of management. New York: Harper College Press; 1977.

[19] Duncombe W, Miner J, Ruggiero J. Empirical evaluation of bureaucratic models of inefficiency. Public Choice 1997;93(1):1–18.

[20] Färe R, Grosskopf S, Lovell CAK. Production frontiers. Cambridge: Cambridge University Press; 1994.

[21] Färe R, Hayes K Grosskopf S, Lundström M and Roos P. Evaluating health care efficiency. In: Blank JLT, Valdmanis V. (Eds.), Evaluating hospital policy and performance: contributions from hospital policy and productivity research, Advances in Health Economics and Health Services Research 18 Chapter 12; 2008, p. 209–28.

[22] Färe R, Grosskopf S, Førsund FR, Hayes K, Heshmati A. Measurement of productivity and quality in non-marketable services: with application to schools. Quality Assurance in Education 2006;14(1):21–36.

[23] Farrell MJ. The measurement of productive efficiency. Journal of the Royal Statistical Society, Series A (General) 1957;120(III):253–81.

[24] Fitz-Gibbon CT, Tymms P. Technical and ethical issues in indicator systems: doing things right and doing wrong things. Education Policy Analysis Archives 2002;10(6):1–26 [Retrieved 01.09.06]⟨http://epaa.asu.edu.epaa/v10n6/⟩.

[25] Fried HO, Lovell CAK, Schmidt SS, editors. The measurement of productive efficiency and productivity growth. Oxford: Oxford University Press; 2008.

[26] Frisch R. Theory of production. Dordrecht: D. Reidel; 1965.

[27] Førsund FR. Measuring efficiency in the public sector. [Memorandum No. 16/2013 from Department of Economics]. University of Oslo; 2013.

[28] Golany B, Tamir E. Evaluating efficiency-effectiveness-equality trade-offs: a data envelopment analysis approach. Management Science 1995;41(7):1172–84.

[29] Golany B, Phillips FY, Rousseau JJ. Models for improved effectiveness based on DEA efficiency results. IIE Transactions 1993;25(6):2–10.

[30] Halme M, Joro T, Korhonen P, Salo S, Wallenius J. A value efficiency approach to incorporating preference information in data envelopment analysis. Management Science 1999;45(1):103–15.

[31] Hammond CJ. Efficiency in the provision of public services: a data envelopment analysis of UK public library systems. Applied Economics 2002;34(5):649–57.

[32] Hanoch G. Homotheticity in Joint Production. Journal of Economic Theory 1970;2(4):423–6.

[33] Hanson T. Efficiency and productivity in the operational units of the armed forces: a Norwegian example. International Journal of Production Economics 2016;179(September):12–23.

[34] Hanson T. Estimating output mix effectiveness: a scenario approach. [Memorandum No 14/2016 from Department of Economics]. University of Oslo; 2016.

[35] Hatry HP. Performance measurement: getting results. 2nd edition 2006. Washington D.C: The Urban Institute Press (; 1999.

[36] Kerstens K. Decomposing technical efficiency and effectiveness of French urban transport. Annales d'Economie et de Statistique 1999;54(April-June):129–55.

[37] Korhonen P, Soisma A, Siljamaki A. On the use of value efficiency analysis and some further developments. Journal of Productivity Analysis 2002;17(1):49–65.

[38] Korhonen P, Tainio R, Wallenius J. Value efficiency of academic research. European Journal of Operations Research 2001;130(1):121–32.

[39] Lancaster KJ. A new approach to consumer theory. Journal of Political Economy 1966;74(2):132–57.

[40] Lauer JA, Lovell CAK, Murray CJL, Evans DB. World health system performance revisited: the impact of varying the relative importance of health system goals. BMC Health Services Research 2004.

[41] Lovell CAK, Walters LC, Wood LL. Stratified models of education production using modified DEA and regression analysis. [Chapter 17]. In: Charnes A, Cooper WW, Lewin A, Seiford L, editors. Data Envelopment Analysis: Theory, Methodology and Applications. London: Kluwer Academic Publishers; 1994. p. 329–51.

[42] Mayston DJ. Analysing the effectiveness of public service producers with endogenous resourcing. Journal of Productivity Analysis 2015;44(1):115–26.

[43] Medina-Borja A, Triantis K. Modeling social services performance: a four-stage DEA approach to evaluate fundraising efficiency, capacity building, service quality, and effectiveness in the nonprofit sector. Annals of Operations Research 2014;221(October):285–307.

[44] Medina-Borja A, Pasupathy KS, Triantis K. Large-scale data envelopment analysis (DEA) implementation: a strategic performance management approach. Journal of the Operational Research Society 2006;58(June):1084–98.

[45] Norwegian Defence Fact and Figures. Oslo: Norwegian Ministry of Defence; 2010.

[46] Ruggiero J. On the measurement of technical efficiency in the public sector. European Journal of Operational Research 1996;90(3):553–65.

[47] Ruggiero J. Efficiency of educational production: an analysis of New York school districts. Review of Economics and Statistics 1996;78(3):499–509.

[48] Ruggiero J, Duncombe W, Miner J. On the measurement and causes of technical inefficiency in local public services: with an application to public education. Journal of Public Administration Research and Theory, J-Part 1995;5(4):403–28.

[49] Schreyer P. Output and outcome – measuring the production of non-market services. Paper Prepared for In: Proceedings of the 30th General Conference of The International Association for Research in Income and Wealth, Portoroz, Slovenia, August 24-30, 2008 ⟨http://www.iariw.org/papers/2008/schreyer.pdf⟩; 2008.

[50] Schinnar AP, Kamis-Gould E, Delucia N, Rothbard AB. Organizational determinants of efficiency and effectiveness in mental health partial care programs. Health Services Research 1990;25(2):387–420.

[51] Shephard RW. Indirect production functions. Meisenheim Am Glan: Verlag Anton Hain; 1974.

Please cite this article as: Førsund FR. Measuring effectiveness of production in the public sector. Omega (2017), http://dx.doi.org/10.1016/j.omega.2016.12.007

[52] Solà M, Prior D. Measuring productivity and quality changes using data envelopments analysis: an application to Catalan hospitals. Financial Accountability Management 2001;17(3):219–45.
[53] Yu M-M. Assessing the technical efficiency, service effectiveness, and technical effectiveness of the world's railways through NDEA analysis. Transportation Research part A 2008;42(10):1238–94.
[54] Yu M-M, Fan C-K. Measuring the performance of multimode bus transit: a mixed structure network DEA model. Transportation Research Part E 2009;45(3):501–15.
[55] Yu M-M, Lin ETJ. Efficiency and effectiveness in railway performance using a multi-activity network DEA model. Omega 2008;36(6):1005–17.

Please cite this article as: Førsund FR. Measuring effectiveness of production in the public sector. Omega (2017), http://dx.doi.org/10.1016/j.omega.2016.12.007