CREPP
Working Papers

Assessing the performance of the public sector[1]

Pierre Pestieau[2]

February 14, 2007

HEC-Management School
University of Liège

CREPP Working Paper 2007/03

Abstract

Amazingly, one is used to hearing harsh statements about inefficient public services. Nor is it surprising to see public sector performance questioned. What is surprising is that what is meant by performance, and how it is measured, does not seem to matter to either the critics or the advocates of the public sector.

The purpose of this paper is to suggest a definition, and a way to measure the performance of the public sector or rather of its main components. Our approach is explicitly rooted in the principles of welfare and production economics. We will proceed in four stages. First of all we present what we call the "performance approach" to the public sector. This concept rests on the principal-agent relation that links a principal, i.e., the State, and an agent, i.e., the person in charge of the public sector unit, and on the definition of performance as the extent to which the agent fulfils the objectives assigned by the principal. The performance is then measured by using the notion of productive efficiency and the "best practice" frontier technique.

In the second stage we move to the issue of measuring the performance of some canonical components of the public sector (education, health care and railways transport), assuming that there is no constraint as to data availability. The idea is to disentangle the usual confusion between conceptual and data problems. In the third stage, we move to real world data problems. The question is then that given the available data, does it make sense to assess and measure the performance of such public sector activities. The final stage is to explain performance or rather lack thereof and to look at the contribution of such an exercise for public policy.

Finally we argue that when the scope is not components but the entirety of the public sector, one should restrict the performance analysis to the outputs and not relate it to inputs.

[1] I am grateful to Tim Coelli, Antonio Estache and Sergio Perelman for their insightful suggestions.
[2] University of Liège, CORE, PSE and CEPR.

Outline

1. Introduction
2. Performance approach
3. Ideal data
4. Real life data
5. Measuring performance without inputs
6. Policy implications and concluding comments

1. Introduction

In both developed and less developed countries, one can speak of a crisis of the public sector. The main charge is that it is costly for what it delivers. Costly at the revenue level (tax distortion, compliance cost) and at the spending level (more could be produced with less). Costly or at least costlier than would be the private sector. Even though this particular charge is rarely supported by hard evidence it has to be taken seriously because of its impact on both policy makers and public opinion.

The purpose of this paper is to address the question of whether we can measure the performance of the public sector, a question that is very general and terribly ambitious. Consequently we will narrow it down by dealing with it in four stages.

In the first stage we consider the public sector as a set of production units including firms, programs, agencies, departments such as social security administration, railways, national health care, education, national defence, … Each unit[3] is supposed to use a number of resources within a particular institutional and geographical setting and to produce a number of outputs, both quantitative and qualitative. Those outputs are related to the objective that have been assigned to the production unit by the principal the authority in charge, i.e., the government.

If the principal were a private firm, the objective assigned to the manager would be simple: maximum profit. However with public authorities one has multiple objectives. For example, in the case of health care or education, maximizing the number of QALYS (years of life adjusted for quality) or the aggregate amount of human capital respectively, is not sufficient. Equity considerations are also among the objectives of health and education policy.

Within such a setting the performance is going to be defined in terms of productive efficiency, and to measure productive efficiency, we will use the efficiency frontier technique. Admittedly productive efficiency is just a part of an overall performance analysis. It has two advantages: it can be measured, and its achievement is a necessary condition for any other type of efficiency. Its main drawback however is that it is based on a comparison among a number of rather similar production units from which a best practice frontier is constructed. Such a comparative approach leads to relative measures, and its quality depends on the quality of the observation units.

We now turn to the second stage of the exercise. We pick up three areas that are typically areas of public spending: health, education, railways transport and we show how performance would be measured if data availability were not a constraint.[4] More precisely, when listing the outputs, the inputs and the environment of our production unit, we assume that the best evidence one can dream of is available. In so doing we avoid one of the classical problems with performance measurement, namely the confusion between conceptual and data problems. In particular at this stage we assume that the quality, not just quantity, of both inputs and outputs can be measured with accuracy.

---

3 In the efficiency frontier literature, one speaks of decision making units (DMU).

4 This is consistent with the Atkinson Review (2005) who insists that the measurement of performance indicators and inputs used has to change according to the type of service or program concerned.

Let us assume that we have good data for the three classes of variables: outputs, inputs, environmental characteristics, and that we have a wide sample of production units covering a certain number of years. We now use the most appropriate method (parametric or not) to measure the best practice frontier, as well as the relative performance of each production unit in each period.

Can we trust these measurements? What can we do with them? If, in the sample that we have, the production units appear to be truly efficient, we can consider that these measurements approximate quite well the theoretical concept of productive (in)efficiency. For an efficient unit cannot produce more with the available resources.

How can we explain the prevailing inefficiency? In the private sector and in a competitive setting, Stigler has doubted the existence of inefficiency. In this paper we deal with the public sector and inefficiency thus can result particularly if there are no efficiency incentive mechanisms. In our "ideal" data we thus have indicators pertaining to the presence of such mechanisms and to the degree of autonomy and competition that our production units (schools, hospitals, railways) are exposed to.

The concept of productive efficiency does not imply a perfect ordering of production units. It just says that for a given vector of inputs and environmental variables one cannot produce more (more of one output, keeping all the others fixed).

Within this "ideal" data, we believe that our measurement of productive inefficiency is quite reliable. However, we rarely have ideal data available. This leads us to the third stage. Under conditions of hard reality the data is insufficient, if not missing. This is even more true for less advanced countries than for OECD member countries. Given the available evidence, the question to be raised is that of whether or not a performance study makes sense. In the literature there are many studies that are basically useless, at least if the motivation is not just to fool people. This occurs because the data is plainly bad or lacking so that one ends up using input variables as performance indicators: number of students for the quality of education or number of hospitals beds in use for the level of health care. We will show for these three activities chosen – health care, education and railways – what has been achieved.

The final stage is to try to explain inefficiency if there is any. A number of factors have been used to explain this: ownership, competition, autonomy, incentive structures, … Whenever the activity of the public sector can be entrusted to the market, privatisation it may be an option even though there is no guarantee that the market always leads to performance improvement. Quite often activities of the public sector cannot be easily privatised, and thus policies to enhance efficiency have to be found within the public sector itself.

The next sections are devoted to illustrating and discussing these stages. An additional section discusses the issue of measuring the performance of the whole public sector without using any input, which applies particularly to the performance of social protection. A final section presents conclusions regarding the merits of the productive efficiency approach, and its expected contributions to public policy.

2. The performance Approach[5]

2.1. Productive efficiency

In this section we present what we call the "*performance approach*" to the public sector. We look at the public sector as a set of departments or programs that can be seen as production units (DMU) with one or several persons in charge that are defined as agents. These agents operate in a given institutional setting; they use a number of inputs and provide a number of outputs that correspond to the objectives assigned to them by the public authority that is defined as the principal. In section 6 we will consider the possibility of measuring the performance of the public sector as a whole, – a more ambitious task but one that has the advantage to including all the transversal spillovers that a sectoral analysis might ignore.

The gist of this paper is that most measures of performance of the public sector are highly questionable. By contrast the productive efficiency criterion appears to be more reliable measurement for appreciating and comparing the performance of the public sector. As already mentioned, our approach rests on the "principal agent" relationship that links the person in charge of the production unit (DMU) with the public authority. Performance is thus defined by the extent to which the agent(s)s fulfil(s) the objectives assigned by the principal.

The principal, that is the State taken as representing Society as a whole, has multiple objectives because of the many dimensions of social welfare. As a result the missions assigned to the agents are also multifold, so that and the performance-assessing issue becomes more complex than that of private firms where the profit level is the performance measurement.

Multidimensional objectives are thus unavoidable when assessing public sector performance. Economists distinguish three main classes of objectives. The first is allocative and it comprises both technical (or productive)[6] and price efficiency. The second one concerns macroeconomic consideration: growth and employment. The third one is one of equity. Equity is a notoriously slippery concept, implying that the incidence of any public activity on income distribution has to be taken into account.

Beyond the recognition that a multidimensional approach is essential, a number of difficulties arise that we will deal with briefly, each in its own turn. First, the objectives assigned to the public sector may not always be completely compatible with one another. One knows, for example, that peak load pricing for the railroad is desirable from an allocative viewpoint, but often distributively objectionable. Thus, when assessing the overall performance of any public activity, a delicate balance must be struck between equity and efficiency considerations.

Second, measuring the degree to which those objectives are satisfied is quite a difficult task. It involves computing first an indicator of partial performance for each of them, and then proceeding to weigh some of those indicators, which cannot be resolved without some basic value judgements.

[5] This section summarizes Marchand *et al.* (1984) and Pestieau and Tulkens (1993).

[6] In the rest of this paper we use "technical" or "productive" efficiency interchangeably; we also use as synonymous the terms "allocative" and "price" efficiency.

Third, the only objective the achievement of which does not impede that of the others is *technical* efficiency. Producing too little or employing too many factors as compared to what is technically feasible cannot be justified in terms of any of the other objectives listed above (macroeconomic, allocative, redistributive). By contrast, hiring labor in (too large) quantities that are allocatively inefficient can be legitimated by macroeconomic considerations of employment policy.

Fourth, the trade-offs between allocative and non-allocative objectives can have effects on the controllability of the production unit. Indeed, it can be allocatively inefficient for two reasons, first because it has to fulfil non-allocative goals and second because the agent is concerned with personal objectives (e.g., the three P's, power, prestige and pay). The difficulty then arises in sorting out these two sources of inefficiency, which has often been presented as an argument in favor of privatisation and deregulation. In a competitive setting a private firm is supposed to be efficient both, technically and allocatively.

In view of the above, we advocate in this paper that the performance of public enterprises be measured and compared on the basis *of productive efficiency* only. Our reasons are twofold: on the one hand, the global performance evaluation problem, that is measuring how close a production unit comes to achieving all the objectives just listed is, in our opinion, too ambitious. Both data and techniques of analysis currently available make such an undertaking unachievable. On the other hand, since productive efficiency allows for evaluations that are consistent with the manifold objectives of the production unit at stake, it definitively constitutes a step in the right direction.

2.2. The frontier method of measuring productive efficiency

Over the course of the last two decades, a number of methods for measuring productive efficiency have been proposed. They all have in common the *frontier* concept which states that efficient units are those operating "on" the frontier of their respective production set, while inefficient firms operate "below" that frontier (i.e. in the interior of their production set). These methods also have in common their reliance on inference. Indeed, statistical data that report on the outputs achieved and the inputs used by a productive organisation do not provide the production set *per se*, no more than they yield the production frontier. Both must be inferred, namely constructed from the data, prior that any efficiency computation can be made. Hence the happy expression of "best-practice frontier" whereby Farrell (1957) originally designated the production frontier that he was the first to derive from statistical data.

To estimate best practice frontiers two main alternative methodologies are available: parametric ones, and non-parametric. The difference between them lies in the technique used to formally describe the frontier. In the first case a usual function with constant parameters – e.g. Cobb-Douglas, or translog – is specified *a priori*. Its parameters are estimated by statistical or other methods in such a way that the graph of the function best "envelops" the data "from above", that is, observations appearing to lie on or below this graph. Then, the efficiency of each observation is computed in terms of the distance between the observation and the graph of the estimated function, now considered to be the frontier of the production set. This distance is usually expressed in terms of the ratio between the achieved output and the output predicted

by the function. Notice that the efficiency measure thus obtained may differ according to which functional form is specified *a priori*. Usage of this parametric method is mostly found in the econometric literature.

In the second case that of the non-parametric methodology, what is specified *a priori* is not an explicit function but rather some formal properties that the points in the production set are assumed to satisfy: e.g. free disposal, convexity (implying either non increasing returns to scale for the frontier, or some form of variable returns) or proportionality (implying constant returns). Data are then "enveloped", too, not by the graph of a function whose parameters are estimated, but instead by determining whether or not each observed point can be considered to be an element of the frontier, under the chosen assumption(s). This is done by solving an appropriately defined system of linear equations – one such system for each observation. The same system can then serve to associate each observation with a numerical efficiency score that measures again the distance between the observation and the frontier. Here as above, the efficiency measure obtained for each observation may differ according to which formal assumptions are specified *a priori* on the production set. Usage of non-parametric methodologies in generally prevalent in the management science and operational research literature. It tends also to prevail in public sector areas in which the concept of production function is not obvious. The two best-known non parametric methods are the DEA for "data envelopment analysis" and FDH for "free disposal hull".

For both categories of methods, parametric and non-parametric, the data set can be either a cross-section of several productive units, or a time series of observations of the same unit. In the first case, direct application of any one of the above methods implicitly assumes that for all units the production set is the same. In the second case, it assumes that the production set remains unchanged over time. Either one of these assumptions can be relaxed, provided information other than just input and output quantities can be included in the analysis. They may include extraneous characteristics specific to some subsets of cross-sectional observations, or time-related characteristics in the case of time-series observations. When time is involved, considerations relating to technical progress (or regress) are particularly relevant, because of the issue of sorting out efficiency gains (movements towards the frontier) from progress (frontier shifts). Both parametric and non-parametric methods have recently offered extensions in this direction.

As both classes of methodologies do operate on exactly the same database, *viz.* input and output quantities, a given data set can always be subjected to both efficiency measurements (and furthermore, within each class, to different functional specifications or set theoretic assumptions). The respective virtues of one or the other approach are, therefore, to be evaluated not so much based on the nature of the data, but rather in terms of the answers each method may provide the analyst in his research.

2.3. Pros and cons

We now turn to the merits of productive efficiency-based indicators of performance as compared to other indicators. As was pointed out above, productive efficiency does not prejudge whether and how other objectives are fulfilled.

A second advantage of the productive efficiency viewpoint is that it relies on physical data that is readily available in many instances, and basically more reliable than financial or accounting data. Third, unlike most partial indicators of performance, productive efficiency can encompass a large number of inputs and outputs, including qualitative aspects without having to go through disputable aggregation. Finally, since the concept of productive efficiency is both intuitive and unambiguous, its measurement generally finds a wide consensus.

Admittedly, production efficiency is only a partial indicator of performance. But compared to traditional indicators, it is by far a more robust method to apply to production units operating in changing market structure or alternative production settings.

The idea that technical efficiency is achievable independently of the other objectives assigned to the production unit and particularly of allocative efficiency has been challenged on various counts. Here we only consider one of the most usual objections that pertains to the term of adjustment. The "short term" objection is quite intuitive and goes as follows: a production unit can be constrained to be technically inefficient if it is forced to employ too much of a factor of production, or to supply too little of a service without being able to quickly adjust its input-output vector so as to remain on the efficiency frontier. Take the example of a public railway company, whose demand just dropped by a lot, and assume that it is not allowed to lay off or reassign any of its employees. It is obvious that in the short run, such a firm is doomed to be technically inefficient. After some time, however it should be able to reduce its idle labor force through attrition, or to use the idle workers to improve the quality or variety of its services.

3. Ideal Data

We now turn to the issue of measuring the productive efficiency of three public activities: railway transportation, secondary education and health care. At the outset, we assume that there are no problems with availability of data. As we are interested in making international comparisons, we assume that we have data for a very large number of countries and also for a long series of years. We will start with railways, clearly the easiest case among the three.

3.1. Railways

As we will see in the next section there are a number of international comparisons of technical efficiency of railways. Here we don't have any restriction on data availability. In particular, we assume that for the output, namely transport of persons and of commodities, we have indicators of quantity but also of quality – comfort, reliability of delivery, punctuality. We can also use data on equity of access to measure accessibility across income scale, time and location. On the input side, we need to know what kind of labor is being used, with sufficient desaggregation concerning skill and experience. We also need information on equipment also disaggregated by type and quality, the length and the quality of the tracks, and finally the different sources of energy. Given that railway companies may operate in very different geographical and institutional environment, it is crucial to have information regarding the management

autonomy, the extent of competition and contestability they are subject to. We also need to know about the constraints regarding price discrimination and community service obligations. All theses items are listed in Table 1 to which we come back when we discuss real life data and recent studies.

In Table 1 we ask for a large number of observations. Ravaillion (2005) insists on this point arguing that the precision of efficiency estimates rises with the number of data points and falls with the number of input and output variables.

Table 1 - Performance measure of national railways

	Ideal data	Data used in recent studies
Outputs	Passenger kilometres	v
	Comfort and punctuality	~
	Freight tons and kilometers	v
	- bulk	~
	- containers	~
	- others	~
	Delivery quality and punctuality	~
	Equity of access	–
Inputs	Labor (disaggregated)	v
	Equipment (disaggregated by type and by quality)	v
	Tracks (length and quality)	~
	Energy (sources)	~
Environment	Geography, stage length	~
	Autonomy	~
	Competition or contestability	~
	Passengers per seat	~
	Price discrimination	~
	Community service obligation	~
Observations	Very large number of years and countries	Too small

Note: v = OK; ~ = more or less; – = unavailable

3.2. Education

We narrow down the study of education to that of secondary public education. We have first to define the objectives of the government (national or local). One can reasonably cite skill in reading and writing, in mathematics and science along with the capacity to eventually find employment or going to college. Given that we deal with people aged 18 who come from different backgrounds, we need indicators regarding average and dispersion. As ideal outputs, we would have:

- individual skills in reading, language, math and science
- individual indicators of employability and access to higher education.

We would use the subgroup of individuals who went through public education. Also these output indicators have to be view incrementally, namely relative to what they were at the start of secondary education.

On the input side one can take two views: physical or financial. The physical inputs are the number and the quality of teachers and administrative staff, the building, and the educational materials. Here too what is important is how inputs are allocated among pupils. Alternatively one can look at overall public spending. Indeed, there are two production processes: the first is from financial spending to physical inputs and the second from inputs to outputs with the possibility of shortcut by going from spending to final outputs.

To compare these input-output vectors and use the traditional measure of best practice, it is important to distinguish comparison within and between countries. In a within country comparison, e.g., among school districts, one has the advantage of dealing with the same institutional and cultural setting. The environmental features that can be useful for performance measurement are the socio-economic characteristics of school districts: educational level, income level and income inequality, unemployment, population size and population density. The role of the family is also very important. In a between-country study, one has to expectedly introduce institutional variables: political decentralization, educational system, mobility of students, share of private schools.

Table 2 presents a list of outputs, inputs and environmental variables that we find ideal. They concern an international comparison, which implies that we need data for several years and a large number of countries.

Table 2 - Performance of public education at the secondary level

	Ideal data	Recent studies
Output	Acquired skills (of sample of 18y. old individuals)	
	- math, science, reading	v
	- foreign languages	–
	Direct employability	–
	Indirect employability (through college)	~
	Happiness	–
	Contribution to R and D	~
Input	Teachers (level and quality)	~
	Staff	~
	Building, equipment	v
	Spatial distribution of schools	–
	Skills at the end of the primary education level	–
Environment	Competition between networks	
	Competition with private schools	~
	Role of the family	–
	Unemployment rate, economic growth	~
	Pedagogical technique	~
Observations	Large number of countries and years	~

Note: see Table 1

3.3. Health care

Since we can ask for any data we wish, we would like to start with data reflecting how the individual expected lifetime and health status have increased following health care use. We would also like to use some information about the quality of the care delivered. With these individual data we can thus measure average values and inequality indicators. Also available would be longevity and quality of health in the absence of public intervention: social insurance, public hospitals and preventive health policy.

The quality of care is important since we are not only interested by the efficiency of medical treatment but also by the way it is delivered.

Turning to the inputs, we would use data on the number of physicians, nurses, staff, hospitals, beds and on the quantity of equipment as well as data on the way all these variables are distributed across space and population. The quality dimension of these inputs is also very important.

Also important is social spending especially if one wants to conduct an intermediate efficiency exercise with social spending as input and the different inputs as outputs.

Table 3 - Performance of public systems of health care

	Ideal data	Current studies
Output	Quality of care (average, distribution)	–
	Incremental life time (average, distribution)	~
	Incremental health status (average, distribution)	~
Input	Physician (speciality, quality, geographical, distribution)	~
	Nurses (speciality, quality, geographical, distribution)	~
	Hospitals (speciality, quality, geographical, distribution)	~
	Social expenditure	v
Environment	Ratio of curative to preventive care	~
	Physical exercise, diet	–
	Age structure	v
	Role of the private sector	~
	Copayment by patients, private insurance	~

What we call the environmental variables is crucial. The effectiveness of healthcare interventions surely influences the extent of premature mortality, but much less than wide social and environmental factors such as sanitation, housing conditions and education, individual lifestyle factors such as smoking, poor diet and lack of physical activity. Intervenes also the age structure of the population. As the number and proportion of older people increases, the prevalence of chronic diseases increases consequently.

4. Real World Data

4.1. Railways

We chose a couple of relatively recent comparative studies, those of Cowie and Reddington (1996) and Land and Lin (2002). They both start with the standard lines against measurements such as profit and labor productivity, and in favor of a production possibility frontier approach. In Cowie and Reddington, devoted to the study of 14 European railways for the year 1992, two alternative outputs are considered: passenger trains kilometers (physical measure) and rail service provision (reflecting public service obligation). On the input side we have the number of employees, state rail capital and population density. The author use parametric (e.g. DOLS) and non parametric (DEA) methods that provide results that are broadly similar. They also compare their results with those of several previous studies and observe several differences. What is striking with this study is that it uses very rough data for both inputs and outputs and none for the institutional setting.

Lan and Lin (2002) cover a wider sample of countries (85) that are much more heterogeneous than those studied above. They also resort to both a parametric (stochastic frontier) and a non-parametric method (DEA). They use length of lines, number of locomotives and cars, and number of employees as input factors and train-kilometer as output. Not surprisingly, they observe that the efficiency scores vary with the method used.

Some earlier studies even though using less complete data attempted to explain efficiency slacks by factors such as managerial freedom [Oum and Yu (1991)] or % of electrified lines [Perelman and Pestieau (1989)].[7]

4.2. Education

In a recent paper Alonso and St. Aubeyn (2005) study the cross-country efficiency of secondary education provision using the PISA indicators for 25 mostly OECD countries. Performance is measured on the basis of these indicators. As inputs, the authors use the ratio of teachers to students and time spent at school. Alonso and St. Aubyn proceed in two stages: first they use DEA (Data Envelopment Analysis) to construct a best practice frontier. Then, they try to explain efficiency slacks with a number of environmental variables. Among these variables, GDP per head and parental educational level happen to be the more significant ones with an obviously positive influence. Pisa indicators concern the skill of 15-year-olds in the following areas: reading, mathematics, problems solving and science literacy. Note that these four indicators are highly correlated (over 0.94).

When discussing the performance of an educational system, e.g. secondary education, there is a wide agreement on the indicators to be used: test scores, earning capacity, and the like. The question is whether the resources devoted to schools influence educational performance. In the US, where research is the most abundant, the evidence is mixed. There is an important school of thought[8] that argues that "the

[7] In the same paper the efficiency measures of postal services are corrected to take into account the role of tendering, which plays a significant role.

[8] Hanushek and Luque (2003), Hanushek (1997).

resources devoted to the schools are not closely or consistently related to the student outcome."

At first sight this result is not surprising. After all the objectives of welfare states regarding education is not merely to maximize the average level of skills and expected earnings. Distribution is also important. Furthermore, schooling is not the only input: family background is also terribly important. As in the case of health where explaining longevity by mere medical spending is ludicrous, explaining educational spending just in terms of educational expenditures is nonsensical.

There is another question that appears when direct quality measures of education are not available. Is it then legitimate to use simple measures of the resources devoted to education as a substitute of performance measure? Fortunately there now exist quality indicators that are comparable: not only PISA but also TIMSS (Third International Mathematics and Science Study). In a recent paper Hanushek and Luque (2003), using this latter performance indicator for 37 countries, show that the influence of resources on student performance appears rather limited, even though it is more positive than in the corresponding analysis of US achievement.

The main lesson to draw from these studies is that it is important to include distributional aspects in the performance indicators. One of the rationales for public education is indeed that it insures every citizen with equal opportunity. In that respect, Grosskopf *et al.* (1997) study the educational performance of Texas school districts by focusing on the conflict between expenditure equality and achievement equality.

It is quite clear that for education there is a long way between the performance indicators found in the literature and what we call the ideal indicators. Too many important variables are missing particularly those pertaining to individual characteristics of teachers and pupils.

4.3. Health care

As in the case of railways and education, we have chosen a couple of recent and typical studies regarding the comparative efficiency of health care systems. The first one is that of Lin *et al.* (2006) who compare health system across Canadian provinces and American States. The second one is by Afonso and St. Aubyn (2004) comparing health expenditure efficiency in OECD countries. The North American study uses DEA to assess the performance of health care at sub-national levels. Inputs are measured alternatively in quantity and in values and they include physicians, nurses, hospitals and pharmaceuticals. Outputs are life expectancy at birth, low birth weight, infant mortality, self-reported health and potential years of life lost following specific diseases. The advantage of using the sub-national approach is that it gives a sample of about 60 DMU from a relatively homogeneous environment. The general conclusion is that most of the Canadian provinces, as well as a fair number of American states are on the efficiency frontier. For the others there is a wide range of efficiency scores. Lin *et al.* (2003) interpret this last result as indicating differences in health policies within the US. Alternative hypotheses such as difference in diet, life style or education are not considered.

Afonso and St. Aubyn (2004) address the same issue for a sample of OECD countries. As inputs, they use in-patients beds, medical technology indicators and employment or total expenditure. As outputs, they use infant mortality and life expectancy. They resort to the two standard non-parametric methods: FDH and DEA.

The average efficiency varies between 0.743 (1 input, 2 outputs and DEA) and 0.959 (3 inputs, 2 outputs and FDH). Korea, Mexico and Turkey are the only countries consistently efficient across the various specifications adopted by the authors. Hence they conclude that: "in less efficient countries there is scope for attaining better results using the very same resources." Such a conclusion would be undisputable if we could be sure that efficiency is being correctly measured.

Finally, let us mention the ranking of the World Health Organization (WHO, 2000) concerning the efficiency of the health care systems of 191 countries. The key health output indicator is the disability adjusted life expectancy. But the WHO also uses a measure of fairness in financing and what is termed "responsiveness". In the overall ranking France gets the top position; Canada is 30th after Morocco, and the US 37th just before Cuba.

These studies are undoubtedly interesting. But their major weakness is that they neglect the fact that what explains most of their performance indicators is not the quantity or the quality of health care interventions, but a set of environmental factors and lifestyle aspects. Smoking, physical inactivity and obesity contribute more to the two main causes of mortality and morbidity, i.e., circulating diseases and cancer, than healthcare. To measure the performance of health care systems, it is thus essential to control for these determinants. This being said, the relationship between these wide determinants and health, though well established, is complex and calls for further research.

Another dimension that is consistently neglected in performance studies is the issue of health inequalities. There exists some work on the redistributive effect of public health care system, but without focus on the productive efficiency issue.[9]

To conclude, we observe a big difference between railways on the one hand and either health and education on the other hand. Railways activities and performance can be accounted for by the inputs we have considered. Health an education even when these are mainly public cannot be explained by just the traditional inputs (labor, capital, etc.), but by a number of environmental factors. Thus, if we could count on data reflecting all the determinants of good education and good health, we could not measure the efficiency of public activities but that of a complex system wherein culture and family are as important as labor and investment.

Does that mean that we cannot assess the performance of, say, the health care system of different countries? As we argue in the next section, we can compare health care systems in terms of their achievements without making any judgment as to their efficiency. To take an example, we can evaluate the performance of students in a class on the basis of their grades. To evaluate them and rank according to their respective merits, committing time, … is much more difficult.

[9] See on this Van Doorselaer *et al.* (1997).

5. Measuring Performance without Inputs

Recently the EU has developed a strategy for improving the performance of social policy among its member states. The central idea is to agree upon a number of social indicators pertaining to poverty, including its intensity and persistence, income inequality, non monetary deprivation, low educational attainment, unemployment, joblessness, poor health, poor housing and homelessness, functional illiteracy and innumeracy, and restricted social participation.[10] We may assume that this selection of indicators is a compromise that reflects the view of the EU country members on what are the objectives of social protection. These indicators are now published on a regular basis with the explicit intent of inducing each country to get as close as possible to the best score. This approach, labelled the *Open Coordination Method*, is in the spirit of the so-called "*yardstick competition*". Accordingly citizens of each country base their electoral choices on the relative performance of their incumbent's policy.

It is tempting to use such indicators to assess the performance of the different European welfare states. Given the small number of the latter, and the big number of the former, it is not surprising that most countries happen to be efficient. However by reducing the number of indicators (using the principal components analysis) and increasing the number of EU countries (to 25) one can hope to get fewer countries on the best practice frontier. One can then raise two interesting questions. What are we measuring? Does it make sense to measure performance without looking at the inputs and, in particular, at the level of spending underlying all the social policies involved.

Our feeling is that to the extent that those indicators have been selected by member countries, some of which have played strategically, picking indicators in areas where they excel, we can consider that a best practice frontier constructed from them is quite meaningful. Naturally the risk of manipulation is unavoidable. A country x, knowing that its strongest point is the relative number of centenarians, will lobby to include this variable among the social indicators. Looking at the indicators chosen, one does not see such an example. One just observes potential conflicts between classic objectives such as increased employment and decreased poverty.

The second question concerns the absence of inputs in the analysis. Even though it does not make economic sense to look at the best practice frontier without taking into account the resources used by the welfare state to achieve its multiple objectives, one realizes that, in general, in popular comparison exercises one rarely looks at the input side. Even in golf the standard ranking does not use the players' handicap. Another reason for not using resources used is that they are much more controversial than the output indicators. To keep the example of social policy, all social expenditures are not used for the purpose at hand and the role of both family (particularly for education and health) and the market are crucial but very difficult to measure. Moreover the government can influence both the market and the family by contributing to social welfare through its tax expenditure policy, i.e., tax break for supplementary insurance, subsidized parental leaves, etc.

[10] Atkinson *et al.* (2002). See also Pestieau (2005).

Without neglecting these observations, using social spending to measure the productive efficiency of welfare states regarding social protection leads to interesting results.[11] Table 4 presents the outcome of a recent study of the performance of European Welfare States using two (poverty, long term unemployment) or four (poverty, long term unemployment, education and health) indicators. The aggregation technique is the DEA that is applied to the case when all countries are assumed to use the same resources and when actual social spending is adopted. It is not surprising to observe that when the number of indicators increase, more countries are on the best practice frontier and it is neither striking to see that with social expenditure used as input more countries are "inefficient" particularly those which are known to have a "generous" welfare states.

Table 4 - Efficiency at getting close to the best practise frontier

	Without Input		With social expenditure to GDP as input	
Countries	POV LTU	POV LTU EDU HEALTH	POV LTU	POV LTU EDU HEALTH
Austria	0.977	1	0.764	0.828
Belgium	0.721	0.925	0.640	0.821
Czech Rep.	1	1	1	1
Denmark	1	1	0.823	0.823
Estonia	0.778	0.419	1	0.609
Finland	0.874	1	0.792	0.929
France	0.733	1	0.533	0.776
Germany	0.826	1	0.600	0.823
Greece	0.453	0.879	0.288	0.635
Hungary	0.936	0.936	0.814	0.814
Ireland	0.942	0.946	1	1
Italy	0.407	1	0.306	0.840
Lithuania	0.789	0.308	0.928	0.382
Netherlands	1	1	0.842	0.870
Poland	0.938	0.452	0.972	0.498
Portugal	0.907	0.907	0.634	0.635
Spain	0.628	0.915	0.548	0.999
United Kingdom	0.930	1	0.641	0.813

Source: Lefèbvre et al. (2005)

Knowing that social spending is not closely related to some indicators our view is that such exercise should be restricted to measuring performance without differential input.

In a widely cited paper, Afonso, Schukrecht and Tanzi (2005) tackle the same problem for 23 OECD member countries. They estimate performance indicators for a number of branches such as health, administration, education, infrastructure, distribution, …

[11] Lefèbvre *et al.* (2005).

In a second stage they use outputs indicators along with public expenditure to measure an overall indicator of technical efficiency. Their main conclusion is that countries with small public sectors tend to be relatively more efficient. Unlike what is suggested above, Afonso *et al.*'s approach does not allow for an easy comparison of performance without and with inputs.

In a sequel of this paper, Afonso *et al.* (2006) apply the same approach to the new member states of the European Union. They show that efficiency is fostered by a number of factors: civil service competence, education and secure property rights.

6. Policy implications and concluding remarks

How can we improve upon the efficiency of the public sector when slacks are observed? To answer such a question, it is important to understand the sources of inefficiency and to distinguish two cases: the case where some other institutional alternative exist and the case where the service or the commodity concerned can only be provided by the public sector.

Starting with the first case, it is widely recognized that what leads to inefficiency is not the type of ownership, but the absence of competition or at least of contestability. This is the view of Leibenstein (1966) who argues that monopolies, public or private, are likely to be what he called X-inefficient. In that vein, some have argued that privatisation without enhancing competition and deregulation could be socially armful. The greatest profit orientation of private ownership may indeed reduce allocative efficiency, but its effect on productive efficiency in ambiguous.

The difficult case is that of activities that can only be undertaken by the State and that furthermore cannot be credibly subject to competition. One thinks for example of departments such as those of justice or social security administration. For these departments, performance incentives should be introduced, which requires obtaining an explicit and visible definition of the task(s) expected from the agents concerned. Transparency and mobility are essential along with a constant questioning of the legitimacy of entitlements.

The gist of this paper is that performance measurement is possible if we can get good data. This is a relatively positive view as opposed to that of those who think that performance measurement in the social area is not possible. Recently, Ravaillion (2005) has questioned the theoretical foundations of measuring and explaining social efficiency. In particular he argues that "social indicators do not stem from anything one could reasonably think as a production function representing a well-defined technology operated by an individual producer with well defined physical inputs." This objection is interesting as it encompasses two important concerns: the limits of aggregation and the quality of data. It becomes overly negative if it rejects any attempt at assessing efficiency. It is clear that efficiency measurement is likely to be more convincing if it focuses on a narrow production unit – e.g. a given school – than on a wide unit – e.g. the whole educational system of a country. At the same time political debate and public policy are often conducted at an aggregate level. Government can be ousted or reforms can be undertaken because the poverty rate is relatively high or because the educational score are catastrophic by comparison with neighbouring countries. The

ideal would be to conduct micro studies to explain observed aggregate inefficiency. This is why we believe that measuring efficiency of health care or education has to go through individuals and focus on the increments due to the action of public spending.

To say that the existing studies are not credible is a bit extreme. What is sure is that they have to be interpreted with caution and that action is needed towards improving the quality and the quantity of data. Economists have to say something about the efficiency of the public sector in times where dismantling it is a hot issue. Otherwise they better change field. This being said, caution is important. What is also important is to try to explain the causes of efficiency slacks.

Our final conclusion is that efficiency per se can only be seriously measured when we deal with production units with a well-defined technology. If we study wide components of the public sector or the whole public sector, then we better restrict ourselves to what we call "measuring performance without inputs."

References

Afonso, A., L. Schuknecht and V. Tanzi (2006), Public sector efficiency. Efficiency for new EU member states and emerging markets, European Central Bank, WP #581.

Afonso A., L. Schuknecht and V. Tanzi (2005), Public sector efficiency: an international comparison, *Public choice*, **123**, 321-47.

Afonso A. and M. St Aubyn (2005), Cross-country efficiency of secondary educations, unpublished.

Afonso A. and M. St Aubyn (2004), Non-parametric approaches to education and health expenditure efficiency in OECD countries, unpublished.

Atkinson, A. (2005), *Atkinson review*: Final Report. Measuring the Government Output and Productivity for National Account, London: Palgrave.

Atkinson, A., B. Cantillon, E. Marlier and B. Nolan (2002), *Social Indicators. The EU and Social Inclusion*, Oxford: Oxford University Press.

Cowie, J. and G. Riddington (1996), Measuring the efficiency of European railways, *Applied Economics*, **28**, 1027-35.

Farrell, M.J. (1957), The measurement of productive efficiency, *Journal of the Royal Statistical Society*, Series A - General, 120, part 3, 253-81.

Gathon, H-J. and P. Pestieau (1995), Decomposing efficiency into its managerial and regulatory components: The case of European railways, *European Journal of Operations Research*, **80(3)**, 500-07.

Grosskopf, S., K. Hayes, L. Taylor and W. Weber (1997), Budget constrained frontier measure of fiscal equality and efficiency in schooling, *Review of Economics and Statistics*, **79(1)**, 116-124.

Hanushek E. and J.A. Luque (2003), Efficiency and equity in schools around the world, *Economics of Education Review*, **23**, 481-502.

Lan L. and T. Lin (2002), Measuring technical and scales efficiency in rail industry: A comparison of 85 railways using DEA and SFA, unpublished.

Lefèbvre, M., S. Perelman and P. Pestieau (2005), Faut-il un ou plusieurs indicateurs d'exclusion sociale?, *Revue belge de Sécurité sociale*, **1**, 59-76.

Lin, Ch., B. Fergusson and A. Laporte (2006), Ranking the health system efficiency among Canadian provinces and American states, unpublished.

Marchand, M., P. Pestieau and H. Tulkens (1984), The performance of public enterprises: Normative, positive and empirical issues, chap. 1, in M. Marchand, P. Pestieau and H. Tulkens (eds), *Concepts and Measurement*, Amsterdam: North Holland.

Oum, T.H. and C. Yu (1991), Economic efficiency of passenger railway systems and implications on public policy, University of British Columbia, unpublished.

Perelman, S. and P. Pestieau (1988), Technical performance in public enterprises: A comparative study of railways and postal services, *European Economic Review*, **32**, 432-41.

Perelman, S. and P. Pestieau (1989), The performance of public enterprises: A comparative efficiency study of railways and postal services, in M. Neumann and K.W. Roskamp (eds), *Public Finance and Performance of Enterprises*, Proceedings of the 43rd Congress of the International Institute of Public Finance, Paris,1987, Detroit, 365-81.

Perelman, S. and P. Pestieau (1993), Comparative performance study of postal services: A productive efficiency approach, *Annales d'Economie et de Statistique*,

Pestieau, P. (2005), *The Welfare State in the EU*, Oxford: Oxford University Press.

Pestieau, P. and H. Tulkens (1993), Assessing and explaining the performance of public enterprise, *FinanzArchiv*, **50**, 293-323.

Ravaillion, M. (2005), On measuring aggregate "social efficiency", *Economic Development and Cultural Change*, **53**, 273-92.

Van Doorslaer, E.K.A., H. Bleichrodt, S. Calonge, U-G. Gertham, M. Gerfin, J. Geurts, L. Gross, U. Hakkinen, R.E. Leu and A. Wagstaff, (1997), Income related inequalities in health: some international comparisons, *Journal of Health Economics*, **16**, 93-112.

WHO (World Health Organisation) (2000), World health report 2000, Geneva, Switzerland <www.int/health-systems-performance>.