Bureau d'économie théorique et appliquée (BETA)
UMR 7522

Documents de travail

« Productivity and performance in the public sector »

Auteurs

Mathieu Lefebvre, Sergio Perelman, Pierre Pestieau

Document de Travail n° 2015 – 15

Juin 2015

Faculté des sciences économiques et de gestion
Pôle européen de gestion et d'économie (PEGE)
61 avenue de la Forêt Noire
F-67085 Strasbourg Cedex

Secrétariat du BETA
Géraldine Del Fabbro
Tél. : (33) 03 68 85 20 69
Fax : (33) 03 68 85 20 70
g.delfabbro @unistra.fr
www.beta-umr7522.fr

Productivity and performance in the public sector

Mathieu Lefebvre, Université de Strasbourg
Sergio Perelman, Université de Liège
Pierre Pestieau, Université de Liège

09 June 2015

Abstract

In times of budgetary difficulties it is not 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 two 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 public authority, 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 to him by the principal. The performance is then measured by using the notion of productive efficiency and the "best practice" frontier technique. We then move to the issue of measuring the performance of some canonical components of the public sector (railways transportation, waste collection, secondary education and health care). We survey some typical studies of efficiency and emphasize the important idea of disentangling conceptual and data problems. This raises the important question that given the available data, does it make sense to assess and measure the performance of such public sector activities?

In the second stage we try to assess the performance of the overall public sector. We argue that for such a level of aggregation one should restrict the performance analysis to the outcomes and not relate it to the resources involved. As an illustration we then turn to an evaluation of the performance of the European welfare states using the DEA approach.

Keywords: performance measure, best practice frontier, social protection.

JEL codes: H50, C14, D24

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 two stages. In the first stage we consider the public sector as a set of production units ranging from the social security administration to public railways. Each unit (one talks of DMU, decision management unit) 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 objectives that have been assigned to the production unit by the principal 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 firms or sectors 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. There exist a large number of efficiency studies concerned with the public sector. Some focus exclusively on public DMU; some others compare public and private DMU. We will present a small sample of these studies. The characteristic of these studies is that even the best of them do not use the ideal data for lack of availability. Particularly missing are the qualitative evidence of both outcomes and inputs. Under the hard reality that data are insufficient, if not missing, the question to be raised is that of whether or not some performance studies make sense.

Whereas in the above studies there is a quite good relation between the outputs and the inputs, when we move up to an aggregate level the link is not anymore clear. For example, public spending in health is not related to the quality of health, for at least two reasons: health depends more on factors such as the living habits or the climate than on spending and health spending can be higher where it is needed, namely in areas of poor health. For this reason, when dealing with the public sector as a whole we prefer to restrict our analysis to the quality of outcomes and not to the more or less efficient relation between resources used and outcomes. The problem becomes one of aggregation of outcome indicators. In this paper, we illustrate our point by evaluating the performance of the European welfare states. We use the DEA technique, which gives different weights to each indicator and each decision unit, here each national welfare state. So doing, we expect that the weight given to a partial indicator to a specific country reflects the importance that this country gives to this indicator. We thus meet the concern of political scientists that different welfare states can have different priorities. This approach has been labeled "benefit of the doubt" by Cherchye et al. (2007) and used by Coelli et al. (2010).

To sum up the spirit of this paper, we believe that the study of the performance of the public sector should comprise two parts: an evaluation of the productive efficiency of its components and an assessment of its achievements as a whole. The next two sections are devoted to these two parts.

2. Efficiency measures of public firms and services

There is a long tradition of efficiency measurement in the public sector and a wide number of studies report the results of performance comparisons concerning public firms and services. As we will illustrate here with some examples – railways transportation, waste collection by municipalities, secondary education and health care – there exists a gap between the ideal data needed for such assessments and the data used in the economic literature. On the one hand, there are the restrictions imposed by data availability – mainly sample size limitations, small number of units and short periods – which constrain the number of dimensions which could be taken into account simultaneously, independently of the methodology used. On the other hand, there is the difficulty to identify and to measure accurately the final outcomes, those which justify the public nature of the firm or the activity and that include quality dimensions. Reliable qualitative information on outputs is not only often missing but also relevant quality features of inputs as well as information on the environmental conditions in which these firms operate are often neglected. We are interested in these deviations from the ideal data. For this purpose, we present a list of variables for each activity under analysis that, in our view, describe what would be the dimensions to be taken into account, assuming no data restrictions.[1]

Furthermore, the objectives assigned by governments and regulatory agencies to public firms and public services are multidimensional. Other than technical efficiency and allocative (price) efficiency, they often include macroeconomic (growth and employment) and distributive (equity) targets. As stated by Pestieau and Tulkens (1993), even if these objectives are not always completely compatible with one another, there is among them one dimension, technical efficiency, which does not impede the achievement of the others. Then, in the survey, which follows, most papers reviewed measure technical efficiency, which is without taking in consideration price, nor cost minimization or profit maximization dimensions. There are however some exceptions.

[1] We rely on Pestieau (2009) for the description of ideal data.

Railways transportation

Our first example is productive efficiency in the sector of public railways transportation. The list of variables presented in Table 2.1 assumes no restrictions in availability of data. Besides output quantities, number and journey length of passenger and tons of freight transported, we include quality indicators: comfort, reliability of delivery and punctuality. Also equity of access is a key dimension: How accessible is railways transportation to different categories of the population, e.g. distinguished by income and by location? Which types of inputs are used in production: i) staff skill and experience; ii) type and quality of equipment; iii) length and quality of tracks; and iv) different sources of energy? In our view, all these dimensions are relevant and would be taken into account in a benchmark study.

Furthermore, given the nature of the activity, railways companies operate, by definition, in different geographical areas and national institutional environments. Therefore, other than geographical characteristics, e.g average stage length and population density, it is crucial to have information on railways sector regulations, e.g. autonomy of management, degree of competition, market contestability. Last but not least, we want to know if they are subject to community service obligations and, perhaps, to constraints regarding price discrimination.

Over the last decade, several papers were published on European railways performance. The aim of most of them was to study the effects of the European Commission railways deregulation policy, launched in the early nineties. The main objectives of this reform, as summarized by Friebel et al. (2008), were: (a) to unbundle infrastructure from operations; (b) to create independent regulatory institutions and (c) to open access to the railways markets for competitors. Most European countries slowly introduced these reforms and this gave the opportunity to proceed to efficiency comparisons among them, particularly between vertically integrated and still unbundled companies. With the exception of Farsi et al. (2005), which study the performance of Swiss regional and local railways networks, and Yu and Lin (2008), which compare European railways performances in 2002, the papers surveyed in Table 2.1 use panel data to draw conclusions in relation with the effect of the ongoing deregulation process in the EU. In Table 2.1, as well as in the following tables in this section, we use different signs to indicate that a particular dimension – output, input or environment (non-discretionary) variable – is taken into account ("+ = yes") or not ("– = no") according to the ideal data, or either if it is taken into account but not completely ("~ = more or less").

INSERT TABLE 2.1

Without doubt, among the potential consequences of the reform, transportation quality and equity of access are key issues, as well as quality of track and of equipment. However, as we learn from Table 2.1, none of these dimensions was taken into consideration in the reviewed studies. Farsi et al. (2005) estimate cost efficiency of 50 subsidized railways in Switzerland using alternative parametric approaches. They show the importance of taking into account firms' heterogeneity in Stochastic Frontier Analysis (SFA). They rely on duality theory and, for this purpose, use input prices instead of quantities ("~" in Table 2.1). Friebel et al. (2008) estimate technical efficiency and productivity growth of 13 European railways using a Linear Structural Relations (LISREL) model. As for the other papers studying European national railways presented here, data for countries with unbundled systems was previously aggregated across all railways companies (infrastructure and operations) operating within a country. Yu and Lin (2008) use a multi-activity Data Envelopment Approach (DEA) approach to compute technical efficiency and effectiveness of 20 European railways in 2002, includng seven Eastern European railways. Growitsch and Wetzel (2009) estimate economies of scope, for integrated vs. unbundled railways companies, using DEA and a data file comprising 27 European Railways over the period 2000-2004. Asmild et al. (2009) address also the effect of reforms on European railways efficiency using a multi-directional DEA approach. This approach allows them to compute, separately, staff and material purchases (OPEX less staff expenditures) cost efficiency and to compare them across Europe taking into account competition and contestability, but also autonomy of companies. Cantos et al. (2010) compute technical efficiency, technical change and productivity growth using DEA, while Cantos et al. (2012) compare DEA and SFA performance measurement. In both cases the authors test the influence of vertical integration vs. unbundled railways controlling simultaneously for population density.

Summing up, none of the studies surveyed here takes into account outputs and inputs quality dimensions. Moreover, none of them control for the potential role on railways outcomes, eventually played, by two institutional particular features: price discrimination and community service obligations.

Waste collection

In most countries around the World, waste collection is a public service whose responsibility falls on local authorities, municipalities in the majority of cases. In Table 2.2 we present the ideal data that should be taken into account in the model. On the output side, we expect to find, besides garbage collected in tons and by type, the service coverage and the quality, as well as scores reflecting environment protection, like the percentage of waste recycling, air and water quality and depletion of non-renewable resources.[2] On the input side, the choice of variables will depend on the DMU characteristics. In the case of public or private firms taking in charge waste collection in one or more municipalities, it would be possible to use physical information on inputs, like labour and equipment, but only if they correspond exactly to the same area for which the outputs are observed. Given the increasing organisational complexity of waste collection, which implies high specialisation and economies of scale, most municipalities outsource these activities. In this case, the input is represented by one variable, the total cost paid by the municipality for waste collection and treatment, which includes direct cost plus outsourcing. Finally, as for the other public services surveyed here, environment (non-discretionary) factors must be taken into consideration. The distance to landfill and the collection frequency are two variables in relation with the geography and population density. Also the age structure and the socio-demographic characteristics of the population must be taken into account, especially when they vary dramatically across municipalities. Moreover, as mentioned before, outsourcing is unavoidable in most cases for municipalities, and is therefore potentially a way to improve the services offered and to

[2] For a detailed presentation of environmental effects of waste collection and treatment, see Emery et al. (2007).

benefit from economies of scale. In the same line of reasoning, the way municipalities price waste collection – weight-based, pay-per-bag, poll tax, …– may influence waste production behavior within the population.

INSERT TABLE 2.2

In Table 2.2 we survey the dimensions taken into account by authors in some recent papers selected here for illustration purposes. Worthington and Dollery (2001) measured cost efficiency in domestic waste management among New South Wales municipalities using DEA. Their study has the particularity to work with a large sample and to take into account the recycling rate and municipalities' geographic and demographic dimensions. García-Sánchez (2008) analysed the performance of waste collection in Spanish municipalities with more than 50,000 inhabitants which "… are obliged by law to provide the same solid waste services…" (p. 329). The authors compute DEA efficiencies using output and inputs quantities and in a second stage test the effect of non-discretionary factors, including socio-economic dimensions. Marques and Simões (2009) study the effect of incentive regulation on the performance of 29 Portuguese waste management operators in 2005. For this purpose, they first compute a two-output (tons collected and tons recycled) two-input (OPEX and CAPEX) DEA model and, in a second stage, analyse the effect of non-discretionary variables, among them the institutional framework (private vs. public and kind of regulation). It is interesting to note that the authors report a detailed list of performance indicators, which includes quality of service and environmental sustainability. This list is published every year by the regulatory agency (Institute for the Regulation of Water and Solid Waste, IRAR), as part of a so-called "sunshine" regulation. This kind of information is also part of our ideal data. Unfortunately, Marques and Simöes decided to not include them in the analysis, because "…they are defined by legislation with high sanctions for non-compliance with laws and regulations" (p. 193).

Finally, we survey in Table 2.2 three recent papers in which the authors study the effect of waste-reducing policies on waste collection and treatment costs of near three-hundred municipalities in Flanders, Belgium. Particularly, these papers distinguish among outputs according to waste types: green, packaging, bulky, residual and EPR (extended producer responsibility: batteries, car tires, electrical equipment …). De Jaeger et al. (2011) compute a DEA model with total costs as input and then test the effect of demographic and socio-economic non-discretionary variables, controlling for institutional differences such as weight based pricing, cooperation agreement and outsourcing. Rogge and De Jaeger (2012) use slightly similar information for a more recent year but rely on a shared input DEA model which allows computing partial cost-efficiency for different waste types. Finally, De Jaeger and Rogge (2013) compute Mamquist productivity indexes for the period 1998 to 2008. The results show that, contrary to expectations, weight-based pricing municipalities did not perform worse in terms of cost efficiency than with pay-per-bag system.

Summing up, recent performance studies on waste collection take into account garbage composition, total costs paid by municipalities and most non-discretionary dimensions. They generally fail to include quality of service and environment sustainability indicators.

Secondary education

In Table 2.3, we illustrate what would be the ideal data to study performances in education and, for this purpose, we choose secondary education. What are the objectives of the government (national or local) on educational matters? One can reasonably expect high skills in reading comprehension, as well as in mathematics, sciences and foreign languages. Given that students come from different backgrounds, we not only need indicators on average scores but also scores' dispersion. On top of that, the capacity to eventually find employment or to access to higher education, going to college, matters too.

On the input side there are two possible views: physical or financial. The physical inputs are the number and the quality of teachers, the administrative staff, the building and the other educational materials. Alternatively one can look at overall public spending. In such case there are two steps embodied: the first step from the financial spending to physical inputs, where inputs prices matter, and the second step from inputs to outputs. Therefore, using financial spending as input implies a potential shortcut and can be a source of bias in performance comparisons. Finally, the skills acquired by students at the end of primary school would be ideally included as input of secondary education.

The environmental variables which must be taken into account vary with the level of aggregation: country, district or school. In a within country comparison one has the advantage of dealing with the same institutional and cultural setting but a number of other dimensions matter, above all the socio-economic environment: income inequality, unemployment, population size and population density. Also the family background and the peer group characteristics are important. In a between-country comparison, one has to expectedly introduce institutional variables like: political decentralization (schools autonomy), competition of private schools, educational system (mobility of students, selectivity, pedagogical techniques …).

In the literature one find best practice comparisons between countries, between districts in a country and between schools, either within or across countries or districts. Most international comparative studies rely on data collected at student level either by the OECD Program for International Student Assessment (PISA) or by the International Association for the Evaluation of Educational Achievement (IEA) Trends in International Mathematics and Science Study (TIMSS).

In Table 2.3 we present the list of outputs, inputs and environmental variables used in a selected number of studies. Afonso and St Aubyn (2006a) use PISA data aggregated by country in international comparisons. As expected, given the small number of observations (25 countries), the number of variables taken into account is reduced to a strict minimum. Sutherland et al. (2009) also compare education efficiency in OECD countries using PISA but relying on disaggregated data at the school level. That allows them to take into account simultaneously the family and socio-economic background, as well as a proxy of capital (computer availability). Both studies also report the results of cost-efficiency comparisons at the national aggregated level using information on educational expenditures. Besides the difficulty to estimate accurately the real cost of education, there is the evidence, from the Hanushek (1997) survey of near 400 studies on US education, that "there is not a strong or consistent relationship between student performance and school resources, at least after variations in family inputs are taken into account" (p. 141). This is not a surprising result given the objectives of welfare states concerning education which are not merely to maximize the average scores and expected earnings, but overall distribution (equity). This is the reason why family background and socioeconomic environment play a key role in many studies.

INSERT TABLE 2.3

In Table 2.3 we report the variables used by Grosskopf et al. (1997) to compare the performance of 310 educational districts in Texas. For this purpose the authors use a parametric indirect distance function approach which takes into account the scores reached by students in previous levels of education. As inputs, other than school teachers, they consider three staff categories: administration, support and teacher aides. Haelermans and De Witte (2012) compare 119 schools performances in the Netherlands looking for the impact of educational innovations. They use a nonparametric conditional (order M) approach which allows controlling for schools heterogeneity, mainly localization. Unfortunately, given data limitations, school inputs are represented by a unique variable: expenses per student. Finally, Wößmann (2003) used probably the largest international data available, 39 countries and more than 260.000 students who participated at the TIMMS study in 1994-95. The author estimates an education production function using parametric models, ordinary and weighted least squares to identify the main drivers of education performances. We choose this study as an illustration to show that ideal data is not an unattainable goal, at least for input and environmental variables. Other than those indicated in Table 2.3, Wößmann (2003) includes several variables controlling for teachers' influence, school responsibility, parents' influence and students' incentives.

To summarize, all these studies use data on students' acquired skills and on the number of teachers, but only in two cases, Grosskopf et al. (1997) and Wößmann (2003), information on output inequality (scores' dispersion), on teachers' quality and, even more important, on students' skills at the end of the primary school. Moreover, none of the studies surveyed in Table 2.3 takes into account information of the course students follow after the end of high school, either the degree of employability or the pursuit of higher education. Such information is obviously difficult to obtain.

Health care

Assuming perfect data availability, we would like to use output data reflecting how the individual expected lifetime and health status increase as a consequence of health care use. At the same time, as indicated in Table 2.4, we would like to consider as output the quality of the care delivered. We are not only interested by the efficiency of medical treatment but also by the way care is delivered. Using individual data it would be possible to compute for these variables average values and inequality indicators (distribution).

On the input side, we would take into account the number and the quality of physicians, nurses and hospitals. The distribution of these inputs, geographical and within the population, matters too. Furthermore, total social spending is a potential substitute of physical and qualitative input variables when the information on inputs is sparse or not reliable.

Potentially environment factors play a crucial role on health care delivered. Other than the age structure of the population, individual lifestyle factors like smoking, poor diet or lack of physical activity matters. Institutions may also have an important role, e.g. the share of prevention in total care expenditures, the importance of the private health sector and of private health insurance, co-payment by patients, etc. Our expectation is that most of the necessary information might be available, even if not in the exactly desired form.

Before turning to a few recent cross-country comparative studies, the first paper in Table 2.4, Crémieux et al. (1999), concerns Canadian provinces and is not interested by the measurement of performance, but by the estimation of an average health care production function. The reason we choose this paper is that it illustrates very well that collecting ideal data is not an impossible task, at least for the ten Canadian provinces over the period 1978-1992. The authors use information on health care outputs and inputs together with detailed information on population socio-economic composition and on individuals' behaviour.

INSERT TABLE 2.4

The other papers surveyed in Table 2.4 deal with cross-country health care data compiled either by the World Health Organisation (WHO) or by the Organization for Economic Co-operation and Development (OECD, 2002). Evans et al. (2000) and Tandon et al. (2001) used as health outputs, respectively, the "disability-adjusted life expectancy" (DALE) measure and a composite measure which takes into account five dimensions: DALE, health inequality, responsiveness-level, responsiveness-distribution and fair-financing.[3] In both studies, based on WHO data on 191 countries over the 1993-1997 period and DEA methodology, two inputs are considered: total health expenditure (public plus private) and average educational attainment in the adult population.

The results of these studies, also reported in The World Health Report 2000 (WHO, 2000), generated some debate and other studies were undertaken using the same WHO data file.[4] Two of them are included in Table 2.4. On the one side, Greene (2004) estimates stochastic frontiers using alternative approaches which take into account countries' heterogeneity and several environmental (non-discretionary) variables, among them income inequality, population density and the percentage of health care paid by the government. On the other side, Lauer et al. (2004) estimated health care systems performance assuming five different outputs, in fact, those included in the composite output measure used by Tandon et al. (2001) but taken separately. The particularity of the DEA approach used by Lauer et al. (2004) is that, rather than considering the five different outputs separately, it assumes an identical (equal to 1.0) input for all countries. It is the so called "benefit of the doubt" model introduced by

[3] For a detailed presentation of these indicators, see Gakidou et al. (2000).
[4] Williams A. (2001) paper generated the debated on WHO (2000) performance measurement.

Melyin and Moesen (1991), which we adopt in the following section to measure the performance of the welfare state in European Union countries.

Finally, we include in Table 2.4 three other studies, Färe et al. (1997), Afonso and St Aubyn (2006) and Joumard et al. (2008), which used OECD data on health care for industrialized countries. Färe et al. (1997) compute productivity Malmquist indexes for 10 countries over the period 1974-1989. The outcome of health care is represented by life expectancy of women at age 40 and the reciprocal of the infant mortality rate. Inputs are the number of physicians and care beds per capita. Afonso and St Aubyn (2006b) computed technical efficiency of 25 countries in 2002 using the free disposable hull (FDH) approach. In their paper the health care production function is specified with two outputs, infant survival rate and life expectancy, and three inputs, the number of doctors, nurses and beds, respectively. In a recent study, Spinks and Hollinsgsworth (2009) recognized that "the OECD health dataset provides one of the best cross-country sources of comparative data available", however they also underline pitfalls in this data, mainly "the lack of an objective measure of quality of life", like additional quality-adjusted life years (QALYs), and a "measure of country-based environmental status". The study by Joumard et al. (2008), appeared as an OECD Economics Department Working Paper, in fact partially answered these criticisms including in the input side a lifestyle variable and a proxy for the economic, social and cultural status of the population. Finally for reasons now discussed the level of aggregation of some of these studies is highly questionable.

Summing up, none of the comparative studies of public health care systems surveyed here takes into account all the output-input dimensions of the ideal data. Moreover, when an output or an input is included, in most cases the authors are obliged to neglect the qualitative and distributional dimensions, given lack of data. And even worse there are the environmental (non-discretionary) factors, in particular data on institutional issues like co-payment by patients, or the ratio of curative to preventive care.

3. The Welfare State Performance in the EU

In the previous section we have seen that many components of the public sector can be submitted to the test of best practices and that such exercise is useful to improve the overall efficiency of the public sector. It is however tempting to try to evaluate the performance of the public sector as a whole. However this raises several questions. In this section we illustrate it by showing estimates of the performance of European public sectors. We have chosen to limit our analysis to that of the welfare state, which is a subset but the most important of the public sector. Two reasons for this: the availability of data and a rather good consensus as to the objectives that the welfare state is supposed to pursue and according to which its performance can be assessed.

The objectives of traditional European welfare states are on the one hand poverty alleviation and inequality reduction and on the other hand protection against life cycle risks such as unemployment, ill health and lack of education. A variety of comparable and regularly updated indicators have been developed for the appraisal of social protection policies in the

28 European Union country members. Here we focus our attention on five of the most commonly used indicators, which concern poverty, inequality, unemployment, education and health. The definitions of the indicators that we use are presented in Table 3.1. The first four indicators, poverty (POV), inequality (INE), unemployment (UNE) and early school leavers (EDU), are such that we want them as low as possible, while life expectancy (EXP) is the only "positive" indicator[5]. The five indicators we are using here cover the most relevant concerns of a modern welfare state and their choice is determined by the objectives of the welfare state. They also reflect aspects that people who want to enlarge the concept of GDP to better measure social welfare generally take into account.[6]

INSERT TABLE 3.1

The values of these indicators for 28 European Union member states are listed in Table A1 in the Appendix for the year 2012.[7] As one sees from Table A1 countries are not good or bad in all respects. Thus we are unable to confidently say that a country *A* is doing better than country *B* unless all five indicators in country *A* are better than (or equal to) those in country *B*. We wish to obtain a performance index of the welfare state, so that we can say that country *A* is actually doing better than country *B*. This is not without making choices regarding the methods we shall use. First the indicators need to be scaled, especially those indicators where a higher value is bad. Second, how should we aggregate the five indicators retained here? Should we use a linear aggregation function or should we rely on more sophisticated techniques as presented above? Third, what about the weights allocated to each of the five indicators in the aggregation process? Furthermore should these weights, if any, vary across countries? Finally, if we assume that these five partial indicators as well as the aggregate indicator measure the actual outcomes of the welfare state (what we call its performance), it would be interesting to also measure the true contribution of the welfare state to that performance and hence to evaluate to what extent the welfare state, with its financial and regulatory means, gets close to the best practice frontier. We argue that this exercise, which in production theory amounts to the measurement of productive efficiency, is highly questionable at this level of aggregation.

When we compare the performance of the welfare state across countries we do not intend to explain it by the social programs comprising the welfare state. We realize that many factors may explain differences in performance. First the welfare state is not restricted to spending but includes also a battery of regulatory measures that contribute to protect people against lifetime risks and to alleviate poverty. Second contextual factors such as family structure, culture and climate, may explain educational or health outcomes as much as anything else.

[5] The data are provided by the EU member states within the OMC (see Eurostat database on Population and Social Conditions). They deal with key dimensions of individual well-being; and are comparable across countries. It is difficult to find better data for the purpose at hand. This being said, we realize that they can be perfected. There is some discontinuity in the series of inequality and poverty indicators. In addition, one could argue that life expectancy in good health is likely to be preferred to life expectancy at birth or an absolute measure of poverty might be better than a relative measure that is too closely related to income inequality. But for the time being, these alternatives do not exist.

[6] See, e.g., the classical measurable economic welfare (MEW) developed by Nordhaus and Tobin (1972) and more recently the Stiglitz report (Stiglitz et al.(2009)).

[7] This section can be viewed as an extension of Coelli et al. (2010) in which we study the performance of social protection in EU15 over the period 1995-2006.

This is why we limit our exercise to what we call performance assessment and argue against the extension to efficiency analysis.

Scaling

The first task is to normalize the five variables in order to make them comparable. Indeed the five indicators listed in Table 3.1 are measured in different units. In the original Human Development Report (HDR, 1990), they use three composite indicators (health, education and income) to derive a Human Development Index (HDI). The authors suggest scaling these indicators so that they lie between 0 and 1, where the bounds are set to reflect minimum and maximum targets. We follow their method and the *n*-th indicator (e.g., life expectancy) of the *i*-th country should be scaled using

$$x_{ni}^{*} = \frac{x_{ni} - \min_{k}\{x_{nk}\}}{\max_{k}\{x_{nk}\} - \min_{k}\{x_{nk}\}}, \tag{1}$$

so that for each indicator the highest score is one and the lowest is zero. For "negative" indicators, such as unemployment, where "more is bad", one specify alternatively:

$$x_{ni}^{*} = \frac{\max_{k}\{x_{nk}\} - x_{ni}}{\max_{k}\{x_{nk}\} - \min_{k}\{x_{nk}\}} \tag{2}$$

so that the country with the lowest rate of unemployment will receive a score of one and the one with the highest rate of unemployment will receive zero. This is obviously not the only way of scaling indicators. Coelli et al. (2010) suggest also other methods.

INSERT TABLE 3.2

Table 3.2 shows the five normalized indicators for our sample of 28 countries. We purposely distinguish between the 15 historical members of the EU (hereafter EU15) and the 13 more recent newcomers (EU13).

Measuring performance

On the basis of the five scaled indicators, we then want to obtain an overall assessment of the welfare state performance. One option would be to follow the HDI method exposed above and calculate the raw arithmetic average of the five indicators. We call it the sum of partial indicators: $SPI_i = \frac{1}{5}\sum_{n=1}^{5} x_{ni}^{*}$. Table 3.3 reports the values as well as the rank of each country. As it appears, we have at the top the Nordic countries, plus Austria, the Netherlands and Luxembourg. But we also have new entrants countries (EU13) doing quite well like Slovenia or Czech Republic which are at the top. At the bottom, we find Latvia, Romania, Bulgaria and Portugal.

However this unweighted summation of partial indicators is quite arbitrary and does not completely respond to the estimation problems we raised earlier. In particular, there is no reason to grant each indicator the same weight. In fact weights could change across indicators and across countries to account for the fact that different countries have different priorities. Indeed some countries may give more weight to employment than to income equality and other countries may give more weight to poverty than to education. One possible solution to this problem is the use of the data envelopment analysis (DEA) method.[8] As seen in the previous section DEA is traditionally used to measure the technical efficiency scores of a sample of firms. In the case of the production of social protection by a welfare state, we could conceptualise a production process where each country is a "firm" which uses government resources to produce social outputs such as reduced unemployment and longer life expectancies. At this stage we will assume that each country has one "government" and hence one unit of input, and it produces the five outputs discussed above.[9]

INSERT TABLE 3.3

The DEA efficiency score are reported on Table 3.3. A number of observations can be made. First, we note that approximately 30% of the sample receives a DEA efficiency score of one (indicating that they are fully efficient). This is not unusual in a DEA analysis where the number of dimensions (variables) is large relative to the number of observations. Second, the mean DEA score is 0.916 versus the mean SPI score of 0.622. The DEA scores tend to be higher because they are relative to observed best practice, while the SPI scores are relative to an "ideal" case where all scaled indicators equal one. Third, the DEA rankings are "broadly similar" to the index number rankings. However a few countries do experience large changes, such as Italy, Spain and Croatia which are ranked 19, 28 and 21 respectively in the index numbers but are found to be fully efficient in the DEA results.[10]

There are two primary reasons why we observe differences between the rankings in DEA versus the SPI index. First, the index numbers allocate an equal weight of 1/5 to each indicator while in the DEA method the weights used can vary across the five indicators. They are determined by the slope of the production possibility frontier that is constructed using the linear programming methods. Second, the implicit weights (or shadow prices) in DEA can also vary from country to country because the slope of the frontier can differ for different output (indicator) mixes.

We can use the shadow price information from the dual DEA linear programming to obtain implicit price weights for each country. The means of these weights are given on Table 3.4. The first thing we note is that the poverty and inequality indicators are given a fairly small weight in the DEA models, while the unemployment indicator is given a weight much larger than 0.3. These results suggest that the uniform weights of 0.2 (used in the SPI) understate the effort needed to improve unemployment (and health and education) outcomes versus reducing inequality and poverty. This may be because unemployment is quite uniformly high

[8] For example, see Coelli et al. (2005) for details of the DEA method. See also Cherchye et al. (2004) who use the DEA in a setting close to this one.

[9] Later in this paper we look at the possibility of measuring the input using government expenditure measures.

[10] The favourable DEA score for Italy and Spain is due primarily to the fact that they have the best life expectancy scores in the sample, which puts them at the edge of the five-dimensional data space and hence gives them a higher likelihood of being found to be efficient because of the convexity of the DEA frontier. Similarly, Croatia has a very good indicator of education.

amongst this group of countries, while inequality levels vary quite a bit, especially when one compares Northern Europe with the rest. Thus, getting a unit change in unemployment outcomes is likely to involve a lot of effort relative to these other indicators.[11]

INSERT TABLE 3.4

Measuring efficiency with or without inputs

In the previous section, we have seen that in traditional measures of production efficiency of public services or public utilities, we gather data on both outputs and inputs to construct a best practice frontier. So doing we are able to say that if a production unit has a certain degree of inefficiency, it means that it can do better with the same quantity of inputs or do as well with less inputs. Even if we showed above that it is difficult to meet the ideal production function, this approach is very useful and has to be used when at least some data are available and there exists of an underlying technology. For example, measuring the efficiency of railways companies with this approach makes sense. Railways transport people and commodities (hopefully with comfort and punctuality) using a certain number of identifiable inputs.

When dealing with the public sector as a whole and more particularly social protection, one can easily identify its missions: social inclusion in terms of housing, education, health, work and consumption. Yet, it is difficult to relate indicators pertaining to these missions (e.g., our five indicators) to specific inputs. A number of papers use social spending as the input, but one has to realize that for most indicators of inclusion, social spending explains little. For example, as we argued earlier, it is well known that for health and education factors such as diet and family support are often just as important as public spending. This does not mean that public spending in health and in education is worth nothing; it just means that it is part of a complex process in which other factors play a crucial and complementary role.

Another reason why using social spending as the input of our 5 indicators is not appropriate comes from the fact that social spending as measured by international organisation is not a good measure of real spending. It does not include subsidies and tax breaks awarded to schemes such as mandatory private pensions or health care and it includes taxes paid on social transfers.[12]

It does not mean that the financing side of the public sector does not matter. It is always important to make sure that wastes are minimized, but wastes cannot be measured at such an aggregate level. It is difficult to think of a well-defined technology, which “produces” social indicators with inputs. To evaluate the efficiency slacks of the public sector, it is desirable to analyse micro-components of the welfare states such as schools, hospitals, public agencies, public institution, railways, etc. such that the studies we presented in the previous section. At

[11] We could also use “weights restricted DEA” (Allen et al., 1997) which allows the weights to be selected within pre-set bounds. This method is a “mix” of these between fixed weights and shadow prices, and is useful if one has strong views regarding the upper and lower bounds that should apply to one or more of these weights.

[12] See Adema et al. (2011).

the macro level, one should stop short of measuring technical inefficiency and restrict oneself to performance ranking.

To use the analogy of a classroom, it makes sense to rank students according to how they perform in a series of exams. Admittedly one can question the quality of tests or the weights used in adding marks from different fields. Yet in general there is little discussion as to the grading of students. At the same time we know that these students may face different “environmental conditions” which can affect their ability to perform. For example, if we have two students ranked number 1 and 2 and if the latter is forced to work at night to help ailing parents or to commute a long way from home, it is possible that he can be considered as more deserving or meritorious than the number 1 whose material and family conditions are ideal. This being said there exists no ranking of students according to merit. The concept of “merit” is indeed too controversial. By the same token, we should not use social spending as an indicator of the “merit” of social protection systems or the public sector as a whole.

4. Conclusions

The purpose of this paper was to present some guidelines as to the question of measuring and assessing the performance of the public sector. We believe that such measurement is unavoidable for two reasons. First, people constantly question the role of the public sector as a whole or of such and such of its components on the basis of questionable indicators. Second, a good measure can induce governments or public firms that are not performing to get closer to the best practice frontier.

We start with the issue of whether or not we have to limit ourselves to a simple performance comparison or we can conduct an efficiency study. We argue that efficiency evaluations can be conducted for components of the public sector when sufficient data are available and there exists a production technology link between resources used and outcomes achieved. When dealing with the overall welfare state or large aggregates such as the health or the education sector we deliberately restrict ourselves to performance comparisons, that is comparisons based only on the outcomes of these sectors. The reason is simple: in those instances, the link between public spending and outcomes is not clear and does not reveal a clear-cut production technology. More concretely, key factors that can affect performance are missing. For example, diet can impact health and family can influence education and yet it is difficult to quantify the role of diet and of family.

We present an overview of recent productive efficiency studies in four areas: railways, waste collection, schools and hospitals. For each of these areas we contrast what we call the ideal set of data with the one that is actually used by researchers. No surprisingly the qualitative data are consistently missing. This weakens the recommendations that can be drawn from these studies and should induce public authorities to further invest in qualitative data collection.

We then turn to the assessment of the performance of 28 European Union country members. The fact that even with a synthetic measure of performance the Nordic countries lead the pack is not surprising. It is neither surprising to see that Mediterranean countries are not doing well. What is surprising is to see that with such a comprehensive concept Anglo-Saxon welfare states do as well as the Continental welfare states such as Germany and France.

References

Adema, W., Fron, P. and M. Ladaique (2011), "Is the European welfare state really more expensive? : Indicators on social spending, 1980-2012; and a Manual to the OECD Social Expenditure Database (SOCX)", *OECD Social, Employment and Migration Working Papers*, No. 124, OECD Publishing.

Afonso A. and M. St Aubyn (2006a), "Cross-country efficiency of secondary education provision: A semi-parametric analysis with non-discretionary inputs", *Economic Modeling*, 23(3), 476-491.

Afonso A. and M. St Aubyn (2006b), "Non-parametric approaches to education and health efficiency in OECD Countries", *Journal of Applied Economics*, 8,2, 227-246.

Allen, R., A.D. Athanassopoulos, R.G. Dyson and E. Thanassoulis (1997), Weight Restrictions and Value Judgements in DEA: Evolution, Development and Future Directions, *Annals of Operations Research*, 73, 13-34.

Asmild, M., Holvad, T., Hougaard, J. L. and D. Kronborg (2009), "Railway Reforms: Do They Influence Operating Efficiency?", *Transportation* 36, 617-638.

Cantos, P., Pastor, J. M. and L. Serrano (2010), "Vertical and Horizontal Separation in the European Railway Sector and its Effects on Productivity", *Journal of Transport Economics and Policy*, 44, 139-160.

Cantos, P., Pastor, J. M. and L. Serrano (2012), "Evaluating European Railway Deregulation Using Different Approaches", *Transport Policy*, 24, 67-72.

Cherchye, L., W. Moesen and T. Van Puyenbroeck (2004), Legitimely diverse, yet comparable: on synthesizing social inclusion performance in the EU, *Journal of Common Market Studies*, **42**(5), 919-955.

Cherchye, L., W. Moesen and T. Van Puyenbroeck (2007), An Introduction to 'Benefit of the Doubt' Composite Indicators Social Indicators Research, 82, 111-145.

Coelli, T., Mathieu Lefebvre and Pierre Pestieau, (2010) "On the Convergence of Social Protection Performance in the European Union," *CESIfo Economic studies*, 56(2), 300-322.

Coelli, T.J., Rao, D.S.P., O'Donnell, C.J. and Battese, G.E. (2005), *An Introduction to Efficiency and Productivity Analysis*, 2nd Edition, Springer, New York.

Crémieux P.-Y., Ouellette P. and C. Pilon (1999), "Health care spending as determinants of health otucomes", *Health Economics*, 8, 627-639.

De Jaeger, Eyckmans, J., Rogge, N. and T. Van Puyembroeck (2011), "Wasteful waste-reducing policies? The impact of waste reduction policy instruments on collection and processing costs of municipal solid waste", *Waste Management*, 31, 1429-1440.

De Jaeger and J., Rogge (2013), "Waste pricing policies and cost-efficiency in municipal waste services: the case of Flanders", *Waste Management & Research*, 31(7), 751–758.

Emery, A., Davies A., Griffiths, A. and K. Williams (2007), "Environmental and economic modelling: A case study of municipal solid waste management scenarios in Wales", *Resources Conservation & Recycling*, 49, 244-263.

Eurostat (2014), *Database on Population and Social Conditions*, http://ec.europa.eu/eurostat/data/database.

Evans D., Tandon A., Murray C. J. L. and A. Lauer (2000), "Comparative efficiency of Nathional Health Systems", WHO GPE Discussion Paper Series 29, Geneva.

Färe, R., Grosskopf, S., Lindgren, B. and J. P. Poullier (1997), "Productivity growth on health-care delivery", Medical Care, 35, 4, 354-366.

Farsi, M., Filippini, M. and W. Greene (2005), "Efficiency Measurement in Network Industries: Application to the Swiss Railways Companies", *Journal of Regulatory Economics*, 28(1), 69-90.

Friebel, G., Ivaldi, M. and C. Vibes (2008), "Railway (De)Regulation: A European Efficiency Comparison, *Economica*", 77, 77-91.

Gakidou, E., Murray, C. J .L. and J. Frenk (2000), "Measuring preferences on health system performance assessment", WHO, EIP/GPE DP 20, Geneva.

García-Sánchez, I. M. (2008), "The performance of Spanish solid waste collection", *Waste Management & Research*, 26, 327-336.

Greene (2004), "Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the World Health Organization's panel data on national health care systems", *Health Economics*, 13, 959–980.

Grosskopf, S., Hayes, K., Taylor, L. L., and W. L. Weber (2001), "On the Determinants of School District Efficiency: Competition and Monitoring", *Journal of Urban Economics*, 49, 453-478.

Grotwisch, C. and H. Wetzel (2009), "Testing for Economies of Scope in European Railways. An Efficiency Analysis", *Journal of Transport Economics and Policy*, 43(1), 1-24.

Haelermans, C. and K. De Witte (2012), "The Role of Innovations in Secondary Shool Performance – Evidence from a Conditional Efficiency Model", *European Journal of Operational Research*, 223, 2, 541-549.

Hanushek, E. (1997), "Assessing the Effects of School Resources on Student Performance: An Update", *Education Evaluation and Policy Analysis*, 19, 2, 141-164.

HDR (1990), *Human Development Report 1990. Concepts and Measurement of Human Development*, United Nations Development Program, Oxford University Press, New York.

Joumard, I., André, C.,Nicq, C. and O. Chatal (2008), "Health status determinants: lifestyle, environment, health care resources and efficiency", OECD Economics Department Working Papers, No. 627, OECD Publishing.

Lauer J. A., Lovell C. A. K., Murray C. J. L. and D. Evans (2004), "World health system performance revisited: the impact of varying the relative importance of health system goals", *BMC Health Services Research*, 4, 19 (doi:10.1186/1472-6963-4-19).

Marques, R. C. and P. Simões (2009), "Incentive regulation and performance measurement of the Portuguese solid waste management services", *Waste Management and Research*, 27, 188-196.

Melyn W. and W. Moesen (1991), "Towards a synthetic indicator of macroeconomic performance: unequal weighting when limited information is available", Public Economics Research Paper No. 17, Katholieke Universiteit Leuven, Leuven, Belgium.

Pestieau, P. (2009), Assessing the Performance of the Public Sector, *Annals of Public and Cooperative Economy*, 80, 133-161, 2009.

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

Rogge, N. and S. De Jaeger (2013), "Measuring and explaining the cost efficiency of municipal solid waste collection and processing services", *Omega*, 41, 653-664.

Spinks, J. and B. Hollinsgsworth (2009), "Cross-country comparisons of technical efficiency of health production: a demonstration of pitfalls", Applied Economics, 41, 417-427.

Stiglitz, J., A. Sen, and J-P. Fitoussi, (2009), Report by the Commission on the Measurement of Economic Performance and Social Progress, <www.stiglitz-sen-fitoussi.fr>

Sutherland, D., Price, R. and F. Gonand (2009), "Improving Public Spending Efficiency in Primary and Secondary Education", *OECD Journal: Economic Studies*, Vol. 2009/1.

Tandon A., Murray C.J. L., Lauer A. and D. Evans (2001), "Measuring overall health system performance for 191 countries", WHO GPE Discussion Paper Series 30, Geneva.

Williams A. (2001), "Science or marketing at WHO? A commentary on World Health 2000". *Health Economics*, 10, 93–100.

World Health Organization (2000), *The World Health Report 2000. Health Systems:Improving Performance World Health Organisation*, Geneva.

Worthington, A. and B. E. Dollery (2001), Measuring efficiency in local government: An analysis of New South Wales municipalities' domestic waste management function", *Policy Studies Journal*, 29, 2, 232-249.

Wößmann, L. (2003), "Schooling Resources, Educational Institutions and Student Performance: The International Evidence", *Oxford Bulletin of Economics and Statistics*, 65, 2, 117-170.

Yu, M. N. and T. J. Lin (2008), "Efficiency and effectiveness in railway performance usinga multi-activity network DEA model", *Omega*, 36, 1005-1017.

Table 2.1: Performance measure of railways activity

	Ideal data	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)
Outputs	Passenger kilometres	+	+	+	+	+	+	+
	Trains comfort and punctuality	–	–	–	–	–	–	–
	Freight tons and kilometres (bulk, containers, …)	+	+	+	+	+	+	+
	Delivery quality and punctuality	–	–	–	–	–	–	–
	Equity of access	–	–	–	–	–	–	–
	Passenger per seat	–	–	+	–	–	+	+
Inputs	Labour (disaggregated)	~	+	+	+	~	+	+
	Equipment (disaggregated by type)	~	–	+	+	~	+	+
	Quality of equipment	–	–	–	–	–	–	–
	Tracks (length)	+	+	+	+	+	+	+
	Quality of tracks	–	–	–	–	–	–	–
	Energy (sources)	~	–	–	–	~	–	–
Environment	Geography, stage length	–	–	+	–	–	+	+
	Autonomy	–	–	–	–	+	–	–
	Competition or contestability	–	+	–	+	+	+	+
	Price discrimination	–	–	–	–	–	–	–
	Community service obligation	–	–	–	–	–	–	–
Observations	Large number companies (countries)	50	12	20	27	23	16	23
	Long period (years)	13	23	1	5	7	20	8

Note: + = yes; ~ = more or less; – = unavailable.

Recent studies: (i) Farsi, Filippini and Greene (2005); (ii) Friebel, Ivaldi and Vibes (2008); (iii) Yu and Lin (2008); (iv) Growitsch and Wetzel (2009); (v) Asmild, Holvad, Hougaard and Kronborg (2009); (vi) Cantos, Pastor and Serrano (2010) and (vii) Cantos, Pastor and Serrano (2012).

Table 2.2: Performance measure of waste collection

	Ideal data	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Outputs	Garbage collected (types and quantities)	~	~	+	+	+	+
	Recycling rate	+	–	+	–	–	–
	Service coverage and quality	~	–	–	–	–	–
	Environment (air and water quality)	–	–	–	–	–	–
Inputs	Labour	–	+	–	–	–	–
	Equipment (disaggregated by type)	–	+	–	–	–	–
	Cost (OPEX and CAPEX)	+	–	+	+	+	+
Environment	Geography (distance landfill, frequency delivery)	+	~	+	–	–	–
	Demography (population density, age)	+	+	+	+	+	–
	Socio-economic characteristics	–	+	+	+	+	–
	Public-private delivery, outsourcing	–	–	+	+	–	–
	Pricing (weight-based, pay-per-bag, …)	–	–	–	+	+	+
Observations	Large number of municipalities (operators)	103	113	29	299	293	272
	Long period (years)	1	1	1	1	1	11

Note: + = yes; ~ = more or less; – = unavailable.
Recent studies: (i) Worthington and Dollery (2001); (ii) García-Sánchez (2008); (iii) Marques and Simões (2009); (iv) De Jaeger et al. (2011); (v) Rogge and De Jaeger (2013); (iv) De Jaeger and Rogge (2013).

Table 2.3: Performance of education at the secondary level

Ideal data		(i)	(ii)	(iii)	(iv)	(v)
Outputs	Acquired skills : reading, maths, science	+	+	+	+	+
	: foreign languages	–	–	–	+	–
	Scores' dispersion	–	–	+	–	+
	Direct employability	–	–	–	–	–
	Indirect employability (through college)	–	–	–	–	–
Inputs	Teachers : number	+	+	+	~	+
	: quality (skills)	–	–	+	–	+
	Administrative staff	–	–	+	~	+
	Building, equipment	–	~	–	~	+
	Skills at the end of the primary education	–	–	+	–	+
Environment	Autonomy / Responsibility	–	–	–	–	+
	Spatial distribution of schools	–	–	–	+	+
	Socio-economic characteristics	–	+	+	–	+
	Family background	+	~	–	–	+
	Unemployment rate, economic growth	~	–	–	~	+
	Pedagogical techniques or innovations	–	–	–	+	+
Observations	Large number of units : countries, districts	25	29	310		39
	: schools				119	n.r.
	Long period (years)	1	1	1	1	1

Notes: + = yes; ~ = more or less; – = unavailable; n.r.=not reported.

Recent studies: (i) Afonso and St Aubyn (2006); (ii) Sutherland, Price and Gonand (2009); (iii) Grosskopf, Hayes, Taylor and Weber (1997); iv) Haelermans and De Witte (2012); v) Bößmann (2003).

Table 2.4: Performance of public systems of health care

Ideal data		(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)
Outputs	Incremental life time (average, distribution)	~	~	~	~	~	~	~	~
	Incremental health status (average, distribution)	~	–	~	~	~	–	–	–
	Quality of care (average, distribution)	–	–	~	~	~	–	–	–
Inputs	Physicians (speciality, quality, distribution)	~	–	–	–	–	~	~	–
	Nurses (speciality, quality, distribution)	–	–	–	–	–	–	~	–
	Hospitals (speciality, quality, distribution)	–	–	–	–	–	~	~	–
	Social expenditure (public and private)	+	~	~	~	–	–	–	~
Environment	Age structure, population density	+	–	–	~	–	–	–	–
	Socio-economic characteristics	+	~	~	~	–	–	–	+
	Individual lifestyle: physical exercise, diet …	+	–	–	–	–	–	–	+
	Ratio of curative to preventive care	–	–	–	–	–	–	–	–
	Role of the private sector	–	–	–	+	–	–	–	–
	Co-payment by patients, private insurance	~	–	–	–	–	–	–	–
Observations	Large number of units : countries, provinces	10	191	191	191	191	10	24	30
	Long period (years)	15	5	5	5	5	15	1	1

Notes: + = yes; ~ = more or less; – = unavailable; n.r.=not reported.

Recent studies: (i) Crémieux, Ouellette and Pilon (1999); (ii) Evans, Tandon, Murray and Lauer (2000); iii) Tandon, Murray, Lauer and Evans (2001); (iv) Greene (2004); (v) Lauer, Lovell, Murray and Evans (2004); (vi) Färe, Grosskopf, Lindgren and Poullier (1997); (vii) Afonso and St Aubyn (2006b); (viii) Joumard, André and Nicq (2010).

Table 3.1: Indicators of Social Protection

	Definition
POV :	At-risk-of-poverty rate after social transfers as defined as the share of persons with an equivalised disposable income below the risk-of-poverty threshold, which is set at 60% of the national median equivalised disposable income (after social transfers).
INE :	Inequality of income distribution as defined as the ratio of total income received by the 20% of the population with the highest income (top quintile) to that received by the 20% of the population with the lowest income (lowest quintile). Income must be understood as equivalised disposable income.
UNE :	Long term unemployed (12 months or longer) as a share of the total active population harmonised with national monthly unemployment estimates.
EDU :	Early school leavers as the percentage of the population aged 18-24 with at most lower secondary education and not in further education or training.
EXP :	Life expectancy as the number of years a person may be expected to live, starting at age 0.

Source: The five indicators are taken from the Eurostat Database on Population and Social Conditions (2014).

Table 3.2 : Normalized indicators and non-weighted sum (SPI) – EU 28, 2012

Country		POV	INE	UNE	EXP	EDU
Austria	AT	0.644	0.789	1.000	0.833	0.836
Belgium	BE	0.600	0.868	0.827	0.762	0.623
Denmark	DK	0.741	0.711	0.925	0.726	0.763
Finland	FI	0.733	0.921	0.962	0.786	0.773
France	FR	0.667	0.711	0.774	0.952	0.643
Germany	DE	0.519	0.763	0.895	0.821	0.691
Greece	EL	0.000	0.158	0.000	0.786	0.652
Ireland	IE	0.585	0.711	0.398	0.810	0.734
Italy	IT	0.274	0.447	0.654	0.988	0.353
Luxembourg	LU	0.593	0.816	0.962	0.881	0.812
Netherlands	NL	0.963	0.947	0.947	0.845	0.778
Portugal	PT	0.385	0.368	0.504	0.774	0.198
Spain	ES	0.067	0.000	0.248	1.000	0.000
Sweden	SE	0.667	0.921	0.970	0.917	0.841
UK	UK	0.511	0.474	0.880	0.821	0.546
Bulgaria	BG	0.141	0.289	0.571	0.036	0.599
Croatia	HR	0.193	0.474	0.308	0.381	1.000
Cyprus	CY	0.622	0.658	0.812	0.833	0.652
Czech R.	CZ	1.000	0.974	0.857	0.476	0.937
Estonia	EE	0.415	0.474	0.669	0.310	0.696
Hungary	HU	0.674	0.842	0.714	0.143	0.647
Latvia	LV	0.289	0.184	0.496	0.000	0.691
Lituania	LT	0.333	0.500	0.586	0.000	0.889
Malta	MT	0.593	0.868	0.857	0.810	0.111
Poland	PL	0.444	0.605	0.774	0.333	0.928
Romania	RO	0.037	0.237	0.842	0.048	0.362
Slovakia	SK	0.733	0.921	0.376	0.262	0.947
Slovenia	SI	0.711	1.000	0.759	0.738	0.990
Mean		0.505	0.630	0.699	0.610	0.668

Table 3.3: DEA performance indicators – EU 28, 2012

Country		SPI	Rank	DEA	rank
			EU15		
Austria	AT	0.821	6	1.000	1
Belgium	BE	0.736	11	0.911	17
Denmark	DK	0.773	8	0.947	14
Finland	FI	0.835	5	0.997	10
France	FR	0.749	9	1.000	1
Germany	DE	0.738	10	0.916	16
Greece	EL	0.319	26	0.846	23
Ireland	IE	0.648	15	0.882	21
Italy	IT	0.543	19	1.000	1
Luxembourg	LU	0.813	7	0.984	11
Netherlands	NL	0.896	1	1.000	1
Portugal	PT	0.446	23	0.799	25
Spain	ES	0.263	28	1.000	1
Sweden	SE	0.863	2	1.000	1
UK	UK	0.646	16	0.904	18
			EU13		
Bulgaria	BG	0.327	25	0.650	28
Croatia	HR	0.471	21	1.000	1
Cyprus	CY	0.716	12	0.902	19
Czech R.	CZ	0.849	3	1.000	1
Estonia	EE	0.513	20	0.758	26
Hungary	HU	0.604	18	0.859	22
Latvia	LV	0.332	24	0.697	27
Lituania	LT	0.462	22	0.896	20
Malta	MT	0.648	14	0.925	15
Poland	PL	0.617	17	0.962	13
Romania	RO	0.305	27	0.842	24
Slovakia	SK	0.648	13	0.965	12
Slovenia	SI	0.840	4	1.000	1
Mean		0.622		0.916	

Table 3.4: Implicit weights - EU 28, 2012

	POV	INE	UNE	EXP	EDU
DEA	0.027	0.160	0.264	0.294	0.255

Appendix

Table A.1: Indicators of social protection

		POV	INE	UNE	EXP	EDU
			EU 15			
Austria	AT	14.4	4.2	1.1	81.1	7.6
Belgium	BE	15.0	3.9	3.4	80.5	12.0
Denmark	DK	13.1	4.5	2.1	80.2	9.1
Finland	FI	13.2	3.7	1.6	80.7	8.9
France	FR	14.1	4.5	4.1	82.1	11.6
Germany	DE	16.1	4.3	2.5	81.0	10.6
Greece	EL	23.1	6.6	14.4	80.7	11.4
Ireland	IE	15.2	4.5	9.1	80.9	9.7
Italy	IT	19.4	5.5	5.7	82.4	17.6
Luxembourg	LU	15.1	4.1	1.6	81.5	8.1
Netherlands	NL	10.1	3.6	1.8	81.2	8.8
Portugal	PT	17.9	5.8	7.7	80.6	20.8
Spain	ES	22.2	7.2	11.1	82.5	24.9
Sweden	SE	14.1	3.7	1.5	81.8	7.5
UK	UK	16.2	5.4	2.7	81.0	13.6
			EU 13			
Bulgaria	BG	21.2	6.1	6.8	74.4	12.5
Croatia	HR	20.5	5.4	10.3	77.3	4.2
Cyprus	CY	14.7	4.7	3.6	81.1	11.4
Czech R.	CZ	9.6	3.5	3.0	78.1	5.5
Estonia	EE	17.5	5.4	5.5	76.7	10.5
Hungary	HU	14.0	4.0	4.9	75.3	11.5
Latvia	LV	19.2	6.5	7.8	74.1	10.6
Lituania	LT	18.6	5.3	6.6	74.1	6.5
Malta	MT	15.1	3.9	3.0	80.9	22.6
Poland	PL	17.1	4.9	4.1	76.9	5.7
Romania	RO	22.6	6.3	3.2	74.5	17.4
Slovakia	SK	13.2	3.7	9.4	76.3	5.3
Slovenia	SI	13.5	3.4	4.3	80.3	4.4