DISCUSSION PAPER SERIES

No. 8096

ON THE WELFARE STATE PERFORMANCE IN THE EUROPEAN UNION

Tim Coelli, Mathieu Lefèbvre, Sergio Perelman and Pierre Pestieau

INTERNATIONAL MACROECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP8096.asp**

ISSN 0265-8003

ON THE WELFARE STATE PERFORMANCE IN THE EUROPEAN UNION

Tim Coelli, CEPA, University of Queensland
Mathieu Lefèbvre, CREPP, University of Liège
Sergio Perelman, CREPP, University of Liège
Pierre Pestieau, CREPP, University of Liège, CORE, CEPR and PSE

Discussion Paper No. 8096
November 2010

Centre for Economic Policy Research
53–56 Gt Sutton St, London EC1V 0DG, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **INTERNATIONAL MACROECONOMICS**. This paper is produced as part of the project Growth and Sustainability Policies for Europe (GRASP), a Collaborative Project funded by the European Commission's Seventh Research Framework Programme, Contract number 244725. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

CEPR Discussion Paper No. 8096

November 2010

ABSTRACT

On the Welfare State Performance in the European Union *

In this paper we use data on five social inclusion indicators (poverty, inequality, unemployment, education and health) to assess and compare the performance of 27 European welfare states (EU27) in 2008. Aggregate measures of performance are obtained using index number methods similar to those employed in the construction of the widely used Human Development Index (HDI). These are compared with alternative measures derived from data envelopment analysis (DEA) methods. We are particularly interested by the comparison between EU15 and the 12 newcomers. As it will appear among the newcomers some countries are ranking among the top performers and others are relegated in the bottom of the ranking.

JEL Classification: H50, C14, D24
Keywords: performance measure, best practice frontier, social protection

Tim Coelli
CEPA, University of Queensland
Brisbane QLD 4072
AUSTRALIA

Email: t.coelli@economics.uq.edu.au

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=172843

Mathieu Lefèbvre
CREPP University of Liege
7 (B31), Boulevard de Rectorat
4000 Liege
BELGIUM

Email: mathieu.lefebvre@ulg.ac.be

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=160602

Sergio Perelman
CREPP University of Liege
7 (B31), Boulevard de Rectorat
4000 Liege
BELGIUM

Email: sergio.perelman@ulg.ac.be

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=116599

Pierre Pestieau
CREPP University of Liege
7 (B31), Boulevard de Rectorat
4000 Liege
BELGIUM

Email: p.pestieau@ulg.ac.be

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=105854

* This paper is produced as part of the project Growth and Sustainability Policies for Europe (GRASP), a Collaborative Project funded by the European Commission's Seventh Research Framework Programme, Contract number 244725.

Submitted 1 November 2010

1. Introduction

In 2004, the European Union then comprising 15 member states (EU15) experienced an enlargement of eight Eastern European countries (Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovakia, and Slovenia), plus the Mediterranean islands of Malta and Cyprus. This was the largest single enlargement in terms of people, landmass and number of countries, though not in terms of GDP. The less developed nature of these countries was of concern to some of the older member states. In 2007 two other Eastern European countries, Romania and Bulgaria, joined the EU now consisting of 27 members (EU27), thus adding to a widespread concern of impoverishment of the Union. In this paper we want to see how these 12 newcomers (EU12) fare in terms of social protection. Are they much different than the older member states? Do they pursue the same objectives of traditional European Welfare states, namely poverty alleviation and protection against life-cycle risks such as unemployment, ill health and lack of education?

Thanks to the Open Method of Coordination (OMC, herafter)[2], a variety of comparable and regularly updated indicators have been developed for the appraisal of social protection policies. For EU15 we already have a series of 12 years but for the newcomers the series is limited to a few years, the most recent. In this paper we focus our attention on five of the most commonly used indicators, which relate to poverty, inequality, unemployment, education and health. 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.

The definitions of the indicators that we use are presented in Table 1. Furthermore, the values of these indicators for 27 European member states are listed in Table A1 in the Appendix for the year 2008. This paper 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. Here we follow the same methodology but because the period is too short we cannot check whether or not performance scores converge. In Coelli et al. we were able to show some catching up.

The key question is that of aggregation. As one sees from Table A1 countries are not good or bad in all respects. In other words, when comparing country *A* with country *B*, we are unable to confidently say that *A* is doing better than *B* unless all five indicators in country *A* are better than (or equal to) those in country *B*. To address this issue we need to use an aggregate indicator of social protection. Perhaps we could use methods similar to those used in constructing the widely used Human Development Indicator (HDI)?[3] That index involves the scaling of its three composite indicators (education, health and income) so that they lie between zero and one, where the bounds are set to reflect

[2] The open method of coordination is a process where explicit, clear and mutually agreed objectives are defined, after which peer review enables Member States to examine and learn from the best practice in Europe. Commonly agreed upon indicators allow each member state to find out where it stands. The exchange of information is designed with the aim of institutionalizing policy mimicking (see Pochet, 2005).

[3] See Anand and Sen (1994).

minimum and maximum targets selected by the authors. The HDI is then constructed as an equal weighted sum of these three scaled indicators.

Table 1: Indicators of Exclusion: - Definitions

	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 Laeken indicators (2007).

In this paper we also wish to construct an aggregate index of social protection, so that we can address questions such as "Is country *A* doing better than country *B*?" Various choices need to be made regarding the methods we use. First, should we use a linear aggregation function as is used in the HDI? Second, how should we scale our indicators – especially those indicators where a higher value is bad (e.g., unemployment)? Third,

should we allocate equal weights to each of the five indicators?[4] If not, how should we determine the weights? Should it be based on a survey of experts, as was done in the World Health Organisation health system efficiency project (see WHO, 2000) or could some form of econometric technique be used? Fourth, should we insist that all countries have the same set of weights or should we allow them to differ so as to reflect different priorities in different countries (for example, see the analysis of the WHO data by Lauer *et al.*, 2004)?[5] Fifth, should we include an input measure, such as government expenditure as a share of GDP on these activities, so as to produce a measure of the efficiency of the social protection system instead of just an output index?

The prime objective of our paper is to go beyond the indeterminacy that is implicit (and voluntarily so) to the OMC and to provide a single index reflecting the performance of European welfare states. Such an index allows us to make performance comparisons across countries and over time.

The question one can raise at this point is that of the relevancy of our partial indicators and thus of our single index as a measure of the performance of the welfare state. This brings us back to the performance approach, according to which the performance of an organisation or of a production unit is defined by the extent to which it achieves the objectives that it is expected to fulfil. In the case of the welfare state, the common view is that it has two main missions: to protect individuals against lifetime risks such as unemployment, sickness, disability, etc. and to alleviate all forms of poverty. Ideally, to check the contribution of the welfare state to the fulfilment of these two missions, one should be able to compute the level of social welfare with and without the welfare state. Namely, with and without the various tax-transfer policies that are part of social protection and the numerous protective regulations of modern welfare states. Needless to say, such an endeavour is, at this point, unrealistic for reasons of methodology and data availability. One thus has to resort to imperfect tools to measure the level of social well-being and the contribution of the welfare state to that level.

The five indicators we are using here cover the most relevant concerns of a modern 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] Their choice is determined by the objectives of the welfare state and, in that respect, they are not as comprehensive as would be considered if one was to attempt to measure social welfare. For example, we do not include a measure of average income or an indicator of environmental quality.

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

[4] The issues of weights and scaling are of course related.
[5] One could also allow the weights to vary across time periods.
[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)).

production theory amounts to the measurement of productive efficiency, is highly questionable at this level of aggregation.

In this paper we focus on the measurement of performance of 27 welfare states in the most recent year. At this point, two words of caution are in order. They concern the scope of our exercise and the quality of data. 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. This is why we limit our exercise to what we call performance assessment and argue against the extension to efficiency analysis.

The second word of caution concerns the data we use. They are provided by the EU member states within the OMC. 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.

The rest of the paper is organised as follows. In the next section we assess the performance of 27 European welfare states for the most recent year, 2008, using a number of social indicators. This involves the construction of an aggregate measure using a similar methodology as that used in the HDI. In section 3 we use a frontier measurement technique known as data envelopment analysis (DEA) to construct an alternative aggregate measure, which allows weights to differ across countries. In section 4 we discuss the issue of performance measurement versus efficiency measurement. A final section provides some concluding comments.

2. Constructing an Aggregate Social Protection Index

We have selected five indicators among those provided by Eurostat. Our selection was based on two concerns: choosing the most relevant data and making sure that they cover the 27 countries. The indicators given in Table 1 reflect different facets of social exclusion.

The five indicators listed in Table 1 are measured in different units. Can we normalize them in such a way that they are comparable? The original Human Development Report (HDR, 1990) suggested that the *n*-th indicator (e.g., life expectancy) of the *i*-th country 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 could alternatively specify:

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

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.

Table 2 shows the normalized partial indicators of social performance for our sample of 27 countries. It is interesting to see the situation of the twelve newcomers compared to the EU15. Though Poland is the best performing country in terms of education, Slovakia and Slovenia performs well in terms of inequality while Czech Republic has also a maximum score for both inequality and poverty. Italy and Denmark are the best in terms of life expectancy and unemployment respectively. Looking at the bottom, Portugal has the worst results for the EU15 but Latvia is doing quite bad for both inequality and poverty.

How can we aggregate these five scaled indicators to obtain an overall assessment of social protection performance? One option is to again follow the HDI method and calculate the raw arithmetic average:[7]

$$SPI_i = \frac{1}{5}\sum_{n=1}^{5} x_{ni}^{*} \ . \tag{3}$$

This has been done and the values obtained are reported in column 7 of Table 2. As it appears, we have at the top the Nordic countries, plus Austria, the Netherlands and Luxembourg. But we also have new entrants countries doing quite well like Slovenia or Czech Republic which are at the top. At the bottom, we find Latvia, Romania, Bulgaria and Portugal.

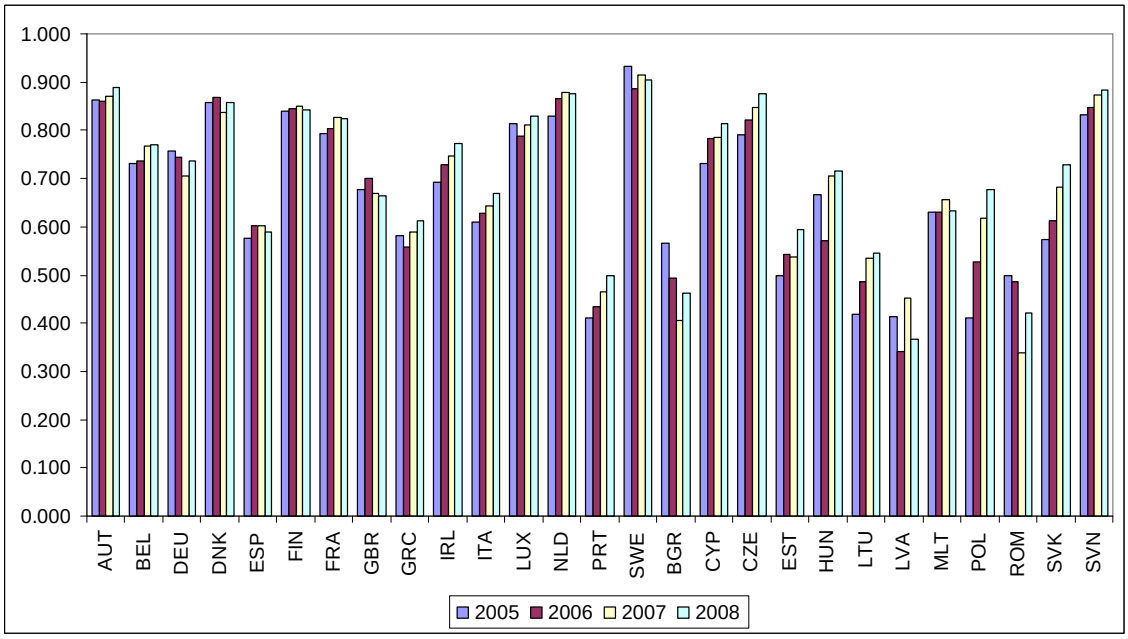
We can also look at the evolution over the period 2005-2008. This is much too short to draw any conclusion as to convergence of performance. On Figure 1, we observe that some countries have decreasing trends in their *SPI* for the 4 years. This is the case of Bulgaria or Romania but also of Germany. Hopefully in 10 years from now robust tests of convergence will be possible.

[7] The acronym, *SPI*, refers to *social protection index.*

Table 2: Normalized Scores and Social Protection Index, EU27 - 2008

	POV	**INE**	**UNE**	**EDU**	**EXP**	*SPI*	Rank
EU15 AUT	0.824	0.923	0.934	0.850	0.896	*0.901*	2
BEL	0.647	0.821	0.541	0.794	0.838	*0.748*	12
DEU	0.647	0.641	0.459	0.800	0.851	*0.688*	14
DNK	0.824	0.949	1.000	0.809	0.704	*0.865*	7
ESP	0.353	0.487	0.754	0.209	0.953	*0.601*	18
FIN	0.706	0.897	0.885	0.859	0.822	*0.866*	5
FRA	0.765	0.795	0.607	0.800	0.976	*0.794*	10
GBR	0.412	0.436	0.852	0.647	0.836	*0.693*	13
GRC	0.353	0.359	0.492	0.712	0.831	*0.598*	19
IRL	0.588	0.718	0.803	0.815	0.824	*0.790*	11
ITA	0.412	0.564	0.574	0.568	1.000	*0.676*	15
LUX	0.765	0.821	0.820	0.753	0.904	*0.824*	8
NLD	0.882	0.846	0.918	0.812	0.887	*0.866*	6
PRT	0.471	0.308	0.475	0.106	0.767	*0.414*	25
SWE	0.824	0.974	0.951	0.821	0.961	*0.927*	1
EU12 BGR	0.294	0.205	0.607	0.712	0.143	*0.417*	24
CYP	0.588	0.821	1.000	0.744	0.911	*0.869*	4
CZE	1.000	1.000	0.721	0.982	0.555	*0.815*	9
EST	0.412	0.590	0.803	0.735	0.240	*0.592*	20
HUN	0.824	0.949	0.492	0.803	0.230	*0.618*	17
LTU	0.353	0.359	0.885	0.929	0.000	*0.543*	23
LVA	0.000	0.000	0.770	0.691	0.053	*0.379*	27
MLT	0.647	0.846	0.672	0.000	0.804	*0.580*	21
POL	0.529	0.564	0.689	1.000	0.381	*0.658*	16
ROM	0.176	0.077	0.689	0.679	0.149	*0.398*	26
SVK	0.882	1.000	0.000	0.971	0.307	*0.569*	22
SVN	0.824	1.000	0.770	0.997	0.743	*0.878*	3
Mean EU 15	0.631	0.703	0.738	0.690	0.870	*0.750*	
Mean EU 12	0.544	0.618	0.675	0.770	0.376	*0.610*	
Mean EU 27	0.593	0.665	0.710	0.726	0.651	*0.688*	

Figure 1: Average indicator SPI 2005-2008



3. Data Envelopment Analysis

The above index construction method described in the previous section uses implicit weights that one could argue are rather arbitrary. One possible solution to this problem is the use of the data envelopment analysis (DEA) method.[8] DEA is traditionally used to measure the technical efficiency scores of a sample of firms. For example, in the case of agriculture, one would collect data on the inputs and outputs of a sample of farms. Output variables could be wheat and beef, while the input variables could be land, labour, capital, materials and services. The DEA method involves running a sequence of linear programs which fit a production frontier surface over the data points, defined by a collection of intersecting hyper-planes. The DEA method produces a technical efficiency score for each firm in the sample. This is a value between zero and one which reflects the degree to which the firm is near the frontier. A value of one indicates that the firm is on the frontier and is fully efficient, while a value of 0.8 indicates that the firm is producing 80% of its potential output given the input vector it has.[9]

In the case of the production of social protection, 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

[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] This is known as an output orientated efficiency score. It reflects the degree to which the output vector of the *i*-th firm can be proportionally expanded (with inputs fixed) while still remaining within the feasible production set defined by the DEA frontier. One can also define input orientated technical efficiency scores, which relate to the degree to which inputs can be contracted (with outputs fixed).

of the paper we will assume that each country has one "government" and hence one unit of input, and it produces the five outputs discussed above.[10]

Table 3: Performance scores and ranks, EU27 - 2008

	SPI		***DEA***		***DEA-I***	
	Scores	**rank**	**Scores**	**rank**	**Scores**	**rank**
EU15 AUT	*0.885*	2	*1.000*	1	*0.770*	20
BEL	*0.728*	12	*0.921*	19	*0.691*	27
DEU	*0.680*	13	*0.931*	17	*0.702*	25
DNK	*0.857*	5	*1.000*	1	*0.712*	23
ESP	*0.551*	21	*0.973*	14	*0.980*	10
FIN	*0.834*	7	*0.978*	13	*0.801*	16
FRA	*0.788*	10	*1.000*	1	*0.739*	21
GBR	*0.637*	15	*0.883*	20	*0.714*	22
GRC	*0.549*	22	*0.866*	21	*0.708*	24
IRL	*0.750*	11	*0.927*	18	*1.000*	1
ITA	*0.623*	18	*1.000*	1	*0.828*	14
LUX	*0.812*	9	*0.938*	16	*0.989*	9
NLD	*0.869*	3	*1.000*	1	*0.823*	15
PRT	*0.425*	24	*0.778*	25	*0.696*	26
SWE	*0.906*	1	*1.000*	1	*0.771*	19
EU12 BGR	*0.392*	25	*0.737*	27	*0.788*	17
CYP	*0.813*	8	*1.000*	1	*1.000*	1
CZE	*0.852*	6	*1.000*	1	*1.000*	1
EST	*0.556*	20	*0.850*	23	*1.000*	1
HUN	*0.659*	14	*0.949*	15	*0.776*	18
LTU	*0.505*	23	*1.000*	1	*1.000*	1
LVA	*0.303*	27	*0.807*	24	*1.000*	1
MLT	*0.594*	19	*0.865*	22	*1.000*	1
POL	*0.633*	16	*1.000*	1	*0.949*	12
ROM	*0.354*	26	*0.755*	26	*0.858*	13
SVK	*0.632*	17	*1.000*	1	*1.000*	1
SVN	*0.867*	4	*1.000*	1	*0.954*	11
Mean EU15	**0.750**		**0.946**		**0.795**	
Mean EU12	**0.610**		**0.914**		**0.944**	
Mean EU27	**0.688**		**0.932**		**0.861**	

The DEA efficiency score are reported in column 4 of Table 3. A number of observations can be made. First, we note that approximately 40% of the sample receives

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

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 0.93 versus the mean SPI score of 0.67. 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 and Lituania which are ranked 18 and 23 respectively in the index numbers but are found to be fully efficient in the DEA results.[11]

Why do we observe differences between the rankings in DEA versus the index numbers? There are two primary reasons. 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 because they are determined by the slope of the production possibility frontier that is constructed using the LP 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.

To investigate this issue, we have used the shadow price information from the dual DEA LP to obtain implicit price weights for each country. The means of these weights are given on Table 4. The first thing we note is that the scaled poverty and inequality indicators are given a fairly small weight in the DEA models, while the health 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 education and health outcomes versus reducing inequality and poverty. This may be because health and reducation outcomes are quite uniformly high 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 health or education outcomes is likely to involve a lot of effort relative to these other indicators.[12]

Table 4: Means of the DEA implicit weights, EU27

	POV	**INE**	**UNE**	**EDU**	**EXP**
EU15	0.035	0.149	0.222	0.291	0.303
EU12	0.000	0.348	0.257	0.307	0.089
EU27	0.043	0.154	0.214	0.276	0.313

[11] The favourable DEA score for Italy is due primarily to the fact that it has the best life expectancy score in the sample, which puts it at the edge of the five-dimensional data space and hence gives it a higher likelihood of being found to be efficient because of the convexity of the DEA frontier. Similarly, Lituania has very good indicators of education and unemployment.

[12] Two weighting methods are described that involve either setting all weights to 0.2, versus using the shadow prices derived from the DEA frontier to set them. A third option is to use "weights restricted DEA" which allows the weights to be selected within pre-set bounds. This method is a "mix" of these two ideas, and is useful if one has strong views regarding the upper and lower bounds that should apply to one or more of these weights. For more on weights restricted DEA methods, see Allen et al (1997).

4. Measuring efficiency with or without inputs

In traditional measures of production efficiency of public services or public utilities, we gather data on both outputs and inputs and construct a best practice frontier using either a parametric (regression) or non-parametric (e.g., DEA) technique. 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. This approach is very useful and should be used to assess the efficiency of the public sector under two key conditions: availability of data and the existence 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[13] use social spending as the input, but one has to realize that for most indicators of inclusion, social spending explains little. For example, 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.

In column 6 of Table 3, we present the DEA measures using social spending as an input (*DEA-I*).[14] The results are not surprising. Countries that spend little and had a low performance now become the most efficient. This is the case of Estonia, Ireland and Latvia. Can we conclude that by spending differently Germany or France would do better? Not necessarily. Doing better can be related to matters independent from social programs: a better diet, a less stressful life, an increased parental investment in education, a more flexible labour market, … For these matters there might be room for public action but not in financial terms.

Does that mean that the financing side does not matter? Not really. It is 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. As a consequence, indicators such as *DEA-I* presented in Table 3, can lead to erroneous conclusions. 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.[15] [16] At the macro level, one should stop short of measuring technical inefficiency and restrict oneself to performance ranking.

[13] Afonso *et al.* (2006, 2005a,b).
[14] See Table A2 in the appendix for data on social expenditure by country in 2008.
[15] For example, see Pestieau and Tulkens (1993).
[16] See Ravaillon (2005) for discussion of this issue.

To again 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.

5. Conclusions

The purpose of this paper was to present some guidelines as to the question of measuring the performance of social protection. We believe that such measurement is unavoidable for two reasons. First, people constantly compare welfare states on the basis of questionable indicators. Second, a good measure can induce national governments that are not well ranked to get closer to the best practice frontier. This is the spirit of the European OMC (Open Method of Coordination) that has lead to the annual publication of indicators of social inclusion for the EU member countries.

In this paper we propose two approaches: one based on a simple average of partial indicators and the other based on Data Envelopment Analysis. The advantage of DEA is to provide flexible and endogenous weights for our inclusion indicators. Another issue we deal with is that of normalization. DEA scores look higher because they are relative to observed best practices and not to a theoretical benchmark like the index numbers.

We then discuss whether or not we have to limit ourselves to a simple performance comparison or we can conduct an efficiency study. Even though we realize that our performance measures depend on the resources invested by the state to finance alternative social protection programs, we deliberately restrict ourselves to performance comparison and argue against the calculation of efficiency measures as it is usually done for micro-units. The reason is simple: the link between public spending and most of our social inclusion indicators is not clear and does not reveal a clear-cut production technology. More concretely, factors that can affect performance are missing. For example, climate can affect health and social attitudes can affect education.

The fact that even with an enlarged measure of social inclusion 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 an enlarged concept Anglo-Saxon welfare states do as well as the Continental welfare states such as Germany and France.

As a final comment, let us come back to the selection of social inclusion indicators. The gist of this paper is to measure the performance of social protection on the basis of its two main objectives: poverty and inequality reduction and protection against lifetime risks. If there were no problem with data availability, the indicators we would like to use would primarily concern the distribution of individual welfare over the lifecycle and across individuals. That ideal measure of welfare would include consumption, education, health and employment. Unfortunately, such evidence does not exist for the EU15 over a sufficiently long period. As a consequence, we have relied upon the indicators made available in the framework of the OMC.

References

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.

Anand, S. and A. Sen (1994), Human Development Index: Methodology and Measurement, *Human Development Report Office Occasional Paper* #12, New York.

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 (2005a), Public sector efficiency: an international comparison, *Public choice*, **123**, 321-47.

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

Charnes, A., Cooper, W.W. and Rhodes, E. (1978), Measuring the Efficiency of Decision Units, *European Journal of Operational Research*, 2(4), 429-444.

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.

Coelli, T.J, M. Lefebvre and P. Pestieau (2010) *CESIfo Economic studies*, 56(2), 300-322.

Coelli, T.J. (1996b), "A Guide to DEAP Version 2.1: A Data Envelopment Analysis (Computer) Program", *CEPA Working Paper* 96/08, Department of Econometrics, University of New England, Armidale.

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.

Cremer, H. and P. Pestieau (2004), Factor mobility and redistribution, in V. Henderson and J. Thisse, eds, *Handbook of Regional and Urban Economics*, vol. 4, North Holland, Amsterdam, 2529-60.

Färe, R., S. Grosskopf, M. Norris and Z. Zhang (1994), Productivity Growth, Technical Progress, and Efficiency Changes in Industrialised Countries, *American Economic Review*, 84, 66-83.

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

Lauer, J. A., Knox Lovell, C.A., Murrauy, C. and Evans, D. (2004), World Health System Performance revisited: the Impact of Varying the Relative Importance of Health System Goals. *BMC Health Service Research*, 4:19.

Lovell and Pastor (1995), Units Invariant and Translation Invariant DEA Models, *Operations Research Letters* 18:3, 147-51.

Nordhaus, W. and J. Tobin (1972), Is growth obsolete? *Economic Growth*, 50th anniversary colloquim, vol. 5, New York, NBER.

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

Pochet, P. (2005), The Open Method of Co-ordination and the Construction of Social Europe. A Historical Perspective, in Zeidlin, J. and Pochet, P. (eds). *The Open Method of Co-ordination in Action. The European Employment and Social Inclusion Strategies*, PIE-Peter Lang, Brussels.

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

Schleifer, A. (1985), A theory of yardstick competition, *Rand Journal of Economics*, **16**, 319-28.

Sinn, H.W., Tax harmonization and tax competition in Europe, *European Economic Review*, 34, 489-504.

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>

Appendix

Table A1: Social protection indicators - 2008

	EDU	INE	EXP	UNE	POV
AUT	10.1	3.7	80.6	0.9	12
BEL	12.0	4.1	80.1	3.3	15
BGR	14.8	6.5	73.3	2.9	21
CYP	13.7	4.1	80.8	0.5	16
CZE	5.6	3.4	77.3	2.2	9
DEU	11.8	4.8	80.2	3.8	15
DNK	11.5	3.6	78.8	0.5	12
ESP	31.9	5.4	81.2	2.0	20
EST	14.0	5.0	74.3	1.7	19
FIN	9.8	3.8	79.9	1.2	14
FRA	11.8	4.2	81.4	2.9	13
GBR	17.0	5.6	80.0	1.4	19
GRC	14.8	5.9	80.0	3.6	20
HUN	11.7	3.6	74.2	3.6	12
IRL	11.3	4.5	79.9	1.7	16
ITA	19.7	5.1	81.6	3.1	19
LTU	7.4	5.9	72.0	1.2	20
LUX	13.4	4.1	80.7	1.6	13
LVA	15.5	7.3	72.5	1.9	26
MLT	39.0	4.0	79.7	2.5	15
NLD	11.4	4.0	80.5	1.0	11
POL	5.0	5.1	75.6	2.4	17
PRT	35.4	6.1	79.4	3.7	18
ROM	15.9	7.0	73.4	2.4	23
SVK	6.0	3.4	74.9	6.6	11
SVN	5.1	3.4	79.1	1.9	12
SWE	11.1	3.5	81.3	0.8	12

Source: Eurostat Laeken Indicators. Income and Living Conditions Database (2010).

Table A2: Social protection spending as a share of GDP - 2008

	% of GDP
AUT	27.1
BEL	28.0
BGR	14.6
CYP	18.1
CZE	18.0
DEU	26.7
DNK	28.1
ESP	20.5
EST	12.3
FIN	24.6
FRA	29.0
GBR	24.8
GRC	23.8
HUN	21.9
IRL	17.6
ITA	25.5
LTU	13.9
LUX	19.0
LVA	10.7
MLT	17.9
NLD	26.8
POL	17.8
PRT	23.4
ROM	12.6
SVK	15.4
SVN	20.8
SWE	29.0