On the Convergence of Social Protection Performance in the European Union.

Tim Coelli,[✓] Mathieu Lefèbvre[†] and Pierre Pestieau[‡]

CREPP Working Paper 2009-03

Abstract

In this paper we use data on five social inclusion indicators (poverty, inequality, unemployment, education and health) to assess and compare the performance of 15 European welfare states (EU15) over a twelve-year period from 1995 to 2006. 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. The influence of methodology choice and the assumptions made in scaling indicators upon the results obtained is illustrated and discussed. We then analyse the evolution of performance over time, finding evidence of some convergence in performance and no sign of social dumping.

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

JEL codes: H50, C14, D24

[✓] CEPA, University of Queensland
[†] CREPP, University of Liège
[‡] CREPP, University of Liège, CORE, CEPR and PSE

1. Introduction

The European Union has adopted an interesting and intriguing approach to achieve some kind of convergence in the field of social inclusion. This approach is known as the "Open Method of Coordination" (OMC).[1] This method requires the definition of common objectives and indicators, which are then used to identify best practice performance. Member states thus regularly know how well they are performing relative to the other states. The implication being, that if a particular state is not performing as well as some other states, it will hopefully be pushed by its citizen-voters to improve its performance.[2]

Thanks to the OMC, a variety of comparable and regularly updated indicators have been developed for the appraisal of social protection policies. 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 definitions of the indicators that we use are presented in Table 1. Furthermore, the values of these indicators for 15 European member states[3] are listed in Table A1 in the Appendix for the 12 year period from 1995 to 2006. If we look closely at the 2006 scores in this table it is evident that some countries do well on some indicators but not on others. For example, Spain has a good health indicator but a very bad poverty indicator, while for Luxembourg it is the converse.

[1] 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).

[2] OMC is related to yardstick competition. See on this Schleifer (1985). Yardstick competition is a method to overcome the information problems or the monitoring restrictions of the authority (here the European Commission). It rests on comparative welfare evaluation. Accordingly, each national government would exert more effort in order to enhance their performance relative to their neighbours. The discipline effect of comparative performance evaluation is expected to generate a sort of "yardstick competition" among national governments, with policicians mimicking the behavior of nearby governments.

[3] These are the 15 European Union members prior to the enlargement of 2005.

Thus, 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 could attempt to construct an aggregate indicator of social protection. Perhaps we could use methods similar to those used in constructing the widely used Human Development Indicator (HDI)?[4] 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 minimum and maximum targets selected by the authors. The HDI is then constructed as an equal weighted sum of these three scaled indicators.

In this paper we 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*?" and "Is country *A* improving over time?" 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?[5] 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)?[6] Fifth, should we include an input measure, such as government expenditure as a share of GDP

[4] See Anand and Sen (1994).
[5] The issues of weights and scaling are of course related.
[6] One could also allow the weights to vary across time periods.

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.[7] 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 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 15 welfare states over a 12 year period. Besides comparing those welfare states, we purport to check if there is any convergence in social inclusion indicators. More importantly, we want to check whether there is any sign of social dumping. Following the increasing integration of European societies, it is feared that social protection might be subject to a “race to the bottom”.[8] As we show convergence is happening and social dumping is not.

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 states and over time or when we check evidence of convergence we do not intend to explain these outcomes by the social programs comprising the welfare state. We realize

---

[7] 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)).

[8] Sinn (1990), Cremer and Pestieau (2004).

that many factors may explain differences in performance or any process of catching up. 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 alleviate poverty.

Second, as we have already noted, 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 (15 here and very soon 27) and over time. 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 due to the transition from ECHP to EUSILC. Also some figures were missing for some years and some countries. For them we filled the gap by simple extrapolation. 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 at least for so many countries and years.

The rest of the paper is organised as follows. In the next section we assess the performance of 15 European welfare states for the most recent year, 2006, using a number of social indicators. This involves the construction of an aggregate measure using a similar methodology to 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, while in section 5 we assess the sensitivity of our results to alternative scaling methods. In section 6 we look at the trend over a period of 12 years, searching for evidence of convergence or divergence, while 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 a sufficient number of years (12) and countries (15). The indicators given in Table 1 reflect different facets of social exclusion. Table 1 provides also the coefficient of correlation among these indicators. The first four indicators poverty (POV), inequality (INE), unemployment (UNE) and education (EDU) are such that we want them as low as possible, while life expectancy (EXP) is the only "positive" indicator.

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 gives the normalized indicators for the year 2006, the most recent for which we have data. For each indicator, the performance of each country can be assessed relative to the best practice (the country with a score of one).

Not surprisingly the Nordic countries lead the pack for inequality, Denmark for unemployment and Finland for education. The Netherlands is first for poverty and Spain for longevity. The worse performers are Portugal for education and inequality, Greece for poverty, Germany for unemployment and Denmark for longevity.

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:[9]

$$SPI1_i = \frac{1}{5}\sum_{n=1}^{5} x_{ni}^{*} \; . \tag{3}$$

This has been done and the values obtained are reported in column 7 of Table 2. As it appears, Sweden is the best ranked and Portugal last. More generally, at the top one finds the Nordic countries, plus Austria, the Netherlands and Luxembourg, and at the bottom the Southern countries.

Given the observed maximum and minimum values in the 2006 data, we can rewrite equation (3) as

[9] The acronym, *SPI*1, refers to *social protection index* number one. The number one is added, because later in this paper we investigate the sensitivity of our results to the use of alternative data scaling methods.

$$SPI1_i = \frac{1}{5}\left[\frac{21 - x_{1i}}{21-10} + \frac{6.8 - x_{2i}}{6.8-3.4} + \frac{5.5 - x_{3i}}{5.5-0.8} + \frac{39.2 - x_{4i}}{39.2-8.3} + \frac{x_{5i} - 78.4}{81.1-78.4}\right] \tag{4}$$

Taking first derivatives with respect to x_{1i} we obtain:

$$\frac{\partial SPI1_i}{\partial x_{1i}} = \frac{1}{5} \times \frac{-1}{21-10} = -0.018\,, \tag{5}$$

and doing the same for the remaining four indicators we obtain –0.059, –0.043, –0.006 and 0.074, respectively.

The ratio of two of these values produces an implicit shadow price ratio

$$\frac{\partial SPI1_i}{\partial x_{ni}} \bigg/ \frac{\partial SPI1_i}{\partial x_{ji}} = \frac{\partial x_{ji}}{\partial x_{ni}}\,. \tag{6}$$

For example, taking poverty and unemployment we obtain –0.043/(–0.018)=2.4. That is, the aggregation process implicitly assumes that reducing the long term unemployment rate by one percent is worth the same as a reduction in the poverty rate of 2.4 percent. Is this what we expected this index to do? What do these relative weights reflect? Are they meant to reflect our social preference function or do they reflect the relative quantities of resources (public expenditure) that would be needed to achieve these things?

To answer these questions we need to do further work. One could perhaps conduct surveys of the general population or of a group of experts to gain some insights into social preferences. However, this exercise is beyond the scope of the current study. Regarding the second option of looking at resource trade-offs, one could attempt to use the sample data to estimate a production technology, and then implicitly use the shadow price information to identify weights. This latter option has the advantage that it can allow weights to differ across countries, depending upon the mix of objectives that a country chooses to focus upon. We investigate the production technology option in the next section.

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.[10] 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.[11]

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

[10] 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. The DEA method is presented in the appendix.

[11] 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.[12]

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 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 0.89 versus the mean SPI score of 0.62. 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 Spain which is ranked 14 in the index numbers but is found to be fully efficient in the DEA results.[13]

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.

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

[13] The favourable DEA scores for Spain are 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.

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 found to be 0.062, 0.067, 0.237, 0.460 and 0.174 for POV, INE, UNE, EDU and EXP, respectively. 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 education indicator is given a weight much larger than 0.2. These results suggest that the uniform weights of 0.2 (used in the SPI) understate the effort needed to improve education outcomes versus reducing inequality and poverty. This may be because education 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 education outcomes is likely to involve a lot of effort relative to these other indicators.[14]

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

[14] 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).

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[15] 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[16]. The results are not surprising. Countries that spend little and had a low performance now become the most efficient. This is the case of Ireland and Luxembourg. 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"

[15] Afonso *et al.* (2006, 2005a,b).

[16] See Table A2 in the appendix for data on social expenditure by country in the period 1995-2006.

social indicators with inputs. As a consequence, indicators such as *DEA1-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.[17] [18] At the macro level, one should stop short of measuring technical inefficiency and restrict oneself to performance ranking.

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. Sensitivity Analysis

In section 2, it was noted that one criticism of the HDI-type approach is that the implicit weights depend upon the composition of the sample. For example, if some of the more recent EU member states were added to the sample we may find that ranges of some

[17] For example, see Pestieau and Tulkens (1993).
[18] See Ravaillon (2005) for discussion of this issue.

indicators may change and hence the relative sizes of the partial derivatives may also change. This could lead to a change in rankings for some countries.

One way to partially, but not fully, address this issue would be to adopt the approach used by Afonso *et al*. (2005) in an international comparison of public sector efficiency. They addressed the scaling issue by scaling each indicator by its sample mean. In the case of "negative" indicators they inverted them before doing this. This method is likely to be more stable because the sample mean is likely to be less sensitive in the face of sample expansion, relative to the sample range (i.e., max−min).

By calculating the means using the 2006 data, we can rewrite equation (3) as

$$SPI2_i = \frac{1}{5}\left[\frac{1}{0.069x_{1i}} + \frac{1}{0.229x_{2i}} + \frac{1}{0.558x_{3i}} + \frac{1}{0.073x_{4i}} + \frac{x_{5i}}{79.9}\right] \tag{10}$$

Taking first derivatives with respect to x_{1n} we obtain:

$$\frac{\partial SPI_{2i}}{\partial x_{1i}} = \frac{1}{5} \times \frac{-1}{0.069(x_{1i})^2} = -2.899(x_{1i})^{-2}. \tag{11}$$

This derivative is not a constant (unlike that in equation 5). It is smaller for larger values of the poverty indicator, *ceteris paribus*. One could argue that this is reasonable since the marginal cost of reducing poverty is likely to be large when poverty rates are very small. However, one could alternatively argue that the social value of reducing poverty in that situation is low.

This derivative when evaluated at the sample mean is equal to −0.012. Furthermore, for the remaining four indicators we obtain −0.042, −0.057, −0.011 and 0.003, respectively. The resulting implicit price ratios are not the same as those obtained using the original method. For example, the poverty and unemployment ratio changes from 2.4 to 4.6.

The results of the two approaches are reported in Table 4 where we see that the choice of indicator does affect rankings for all but five countries (Belgium, Spain, Italy, Luxembourg and Portugal). Most movements are small, although France and Denmark move by four and five places respectively, which is not insignificant in a table of 15 countries. We also note that the mean score SPI2 is higher, at one. This is not unexpected, since the average indicator in this case is one while in the previous case the *maximum* was one.

Also reported in Table 4 are a third set of results, *SPI*3. These are derived using a method closely related to the HDI approach. The only difference is that instead of using the sample minimum and maximums, alternative "goalposts" are used, following the suggestion provided in Anand and Sen (1994). In that paper, the authors note that using in the original HDR (1990) minimum and maximum sample values in the scaling process will be problematic when between year comparisons are made because the minimum and maximum sample values will differ from year to year. They instead suggested the use of "goalpost" values, which reflect their assessments of retrospective and prospective limits. For example, they suggest a range of 35 to 85 for life expectancy and 0 to 100 for education. Using similar logic to theirs we could argue that the ranges for poverty and unemployment should also be 0 to 100. Identifying a range for the inequality indicator is more difficult. Hence we have decided to invert it and multiply it by 100, meaning that it now has a natural range from 0 (the poorest 20% earn nothing) to 100 (the poorest 20% earn the same amount as the richest 20%).

The *SPI*3 results are reported in Table 4. The ranks are similar to *SPI*1, though some countries have a notable change in rank, with Austria, Denmark and Finland all improving by three or more places. We also observe that the mean score is higher and the range of scores is narrower, ranging from 0.68 to 0.79, as compared with 0.16 to 0.91 for *SPI*1. This is again as expected, since the "goalposts" for each of the five original indicators are wider than the sample ranges.

In Table 8, we give the correlation coefficients for several measures. The correlations between the three alternative indices are all 88% or higher, indicating strong but not perfect correlation between these indices.

DEA Analysis

The above two alternative sets of scaled indicators were also used in DEA models. The results are reported in Table 5, along with the original set of scores. The first point to note is that the mean DEA score increases from 89% for DEA1 to 99% for DEA2 and DEA3. This is purely a consequence of the different scaling methods used, and emphasises that when data does not have a natural scale, one should take great care in interpreting the relative sizes of efficiency scores.[19]

The rankings in the three different sets of DEA results do vary to some extent, with a few countries, such as the UK, experiencing some large changes. Overall, the DEA rankings appear to be more stable than the SPI rankings. This is most likely due to the fact that the DEA implicit weights can self-adjust to the different scaling methods, while the SPI measures have fixed rigid weights.[20]

The means of the implicit weights from the three DEA models are listed in Table 7. The weights change notably across the three models. In particular, the weights in the *DEA2* model vary notably from 0.2, with the life expectancy indicator given a large

[19] Unfortunately, the invariance properties of DEA models are not widely recognized. Most standard DEA models are invariant to multiplicative scaling but they are generally not invariant to additive translation or nonlinear transformations, such as inversion. See Lovell and Pastor (1995) for a detailed discussion of scaling and translation invariance properties in DEA models.

[20] The results for the DEA models with social expenditure as an input are reported in Table 6 for completeness. The discussion of these results would be similar to that associated with Table 5.

weight of in excess of 0.7. This is likely to be a consequence of the fact that it is the only indicator that was not inverted prior to inclusion in the DEA model. This observation should serve as a warning to others who may apply data transformations to indicators prior to including them in an equal-weighted aggregate index calculation. The choice of what transformation to use (in this case inversion version linear transformation) can have a substantive effect upon the results obtained.

In Table 8 we provide sample correlations across our 6 indices/scores. One observes reasonably strong correlations between the various measures, which is reassuring. Thus, in section 6, when we study the evolution of performance over a 12-year period, we will focus our attention on one set of indicators: *DEA1* and *SPI1*, without the risk of our choice having a large effect on our results.

6. Convergence

Thus far, we have focused on the year 2006. We now use data on five social indicators covering 12 years. It is interesting to see whether or not we observe any trend and particularly any convergence. In other words, do we see that countries that did not fare well at the beginning of this twelve-year period do progressively catch up? To study that evolution, we use our two approaches: average indicator and DEA, but we restrict the analysis to the HDI normalization.

For the average indicator *SPI1*, we have normalized the primary indicators over the whole period. In other words a value of 1 is given to the country and the year that has the best indicator (e.g., the lowest poverty rate) and vice-versa for the value of 0. These normalized indicators are listed in Table A1 in the appendix. Consider the poverty indicator. With the lowest poverty rate we have Sweden in 1995-1999 and Finland in 1995-1997. Their normalized indicator is thus 1. The highest poverty rate is in Portugal in 1995. Summing up these normalized indicators and dividing by 5, we obtain an average indicator for each country and each year. These are presented in Table 9 and Figure 2.

In Figure 2 it is evident that in all countries (except Sweden) there has been a sharp improvement, particularly among the lagging countries: Spain, Ireland and Portugal. This seems to indicate some catching up with the best student of the "European class", namely Sweden. To check whether there is convergence, one can regress the variation in the indicator at hand, here *SPI1*, against its value in 1995. The results of this regression are presented in Figure 3. As we can see, with a slope coefficient of -0.109 and a R^2 of 0.9, we have clear convergence.[21]

DEA technical efficiency measures for each year are reported in Table 9. Here too we can see that many countries with a score below 1 improve over the 12-year period. However we have to keep in mind that these DEA technical efficiency measures are relative to a best practice frontier that is constructed using data only from the year at hand. Hence, movements in this frontier from year to year are not captured by the technical efficiency measure.

In other words, the performance of a country over time can be decomposed in two elements. Take two countries *A* and *B*, and two years. *A* is on the frontier in the two years, but it is doing better from one year to the other, which means that the frontier moves up. We look at the performance of *B* with respect to that moving best practice frontier; we can decompose it into (i) the change in distance with respect to the best practice frontier, (ii) the change of the best practice frontier itself. Table 9 is only concerned with the first change.

[21] For the SPI and the DEA we have tested the case of convergence for the 3 types of scaling. However we only report here the results pertaining to the first type. The other results are available on request.

To accommodate the two types of changes, we use a technique that is used in production theory. It rests on the Malmquist index that gives the rate at which the frontier moves up and the rate at which the distance to the frontier changes over time.[22] Table 10 gives the yearly changes and the average change. The countries with the lowest average increase are the three Nordic countries that are also those with the highest levels but also Portugal.

The indicators presented on Table 10 can be decomposed in a change in the frontier (Technical change) and a change in the distance to the frontier (Efficiency change).[23] Those two components are given in Table A3 in the appendix.

As with the indicator *SP11*, we wish to check whether or not there is some catching up with our *DEA1* measure. In Figure 4 we regress the average annual Malmquist TFP growth measure against the DEA1 measure in 1995. As we can see, there is convergence with a $R^2 = 0.36$. When we only consider the variation in "technical efficiency" the convergence appears to be stronger with a $R^2 = 0.55$ as it appears on Figure 5. This seems to imply that relative to their own best practice frontier, European countries tend to converge unambiguously.

7. 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

[22] See Coelli et al (2005) for details.

[23] The formula is given by Malmquist + 1 = (efficiency change + 1) * (technical change + 1).

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. In our sensitivity analysis, we show that our results are somehow sensitive to the scaling indicators. We consider three types of scaling and do not have solid grounds to prefer one over the other. However, they fortunately lead to quite similar evaluations.

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 two questions: (i) Do we have to limit ourselves to a simple performance comparison or can we conduct an efficiency study? (ii) How do our performance measures evolves over time? Do we witness any race to the bottom? 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.

Another finding of our paper is that there appears to be some clear convergence in performance among European countries, suggesting that the Open Method of Coordination may be achieving its desired outcome. This latter result is quite interesting. There is so much talk of social dumping and of a race to the bottom that it is comforting to realize that most countries perform better and in a converging way.

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. (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>

Table 1: Indicators of Exclusion: - Definitions and Correlations

	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.

	Correlation				
	POV	INE	UNE	EDU	EXP
POV	1.000				
INE	0.908	1.000			
UNE	0.397	0.390	1.000		
EDU	0.647	0.774	0.272	1.000	
EXP	-0.048	-0.085	0.014	-0.209	1.000

Source: The five indicators are taken from the Eurostat database on Laeken indicators (2007).

Table 2: Normalized Scores and Social Protection Index, 2006

	POV	INE	UNE	EDU	EXP	*SPI1*	Rank
AT	0.73	0.91	0.89	0.96	0.63	0.82	4
BE	0.55	0.76	0.28	0.86	0.41	0.57	9
DE	0.73	0.79	0.00	0.82	0.56	0.58	8
DK	0.82	1.00	1.00	0.92	0.00	0.75	6
ES	0.09	0.44	0.15	0.30	1.00	0.40	14
FI	0.73	0.94	0.77	1.00	0.44	0.78	5
FR	0.73	0.82	0.79	0.87	0.93	0.83	2
GR	0.00	0.21	0.87	0.75	0.41	0.45	12
IE	0.27	0.56	0.57	0.87	0.48	0.55	10
IT	0.09	0.38	0.34	0.60	0.63	0.41	13
LU	0.64	0.76	0.87	0.71	0.37	0.67	7
NL	1.00	0.88	0.81	0.85	0.59	0.83	2
PT	0.27	0.00	0.36	0.00	0.19	0.16	15
SE	0.82	0.97	0.94	0.88	0.96	0.91	1
UK	0.18	0.41	0.91	0.85	0.33	0.54	11
Mean	**0.51**	**0.66**	**0.64**	**0.75**	**0.53**	**0.62**	

Table 3: Performance scores and ranks, 2006

	SPI1		***DEA1***		***DEA1-I***	
	Scores	**rank**	**Scores**	**rank**	**Scores**	**rank**
AT	0.82	4	1.000	1	0.972	8
BE	0.57	9	0.866	13	0.744	13
DE	0.58	8	0.879	12	0.872	11
DK	0.75	6	1.000	1	0.946	9
ES	0.40	14	1.000	1	1.000	1
FI	0.78	5	1.000	1	1.000	1
FR	0.83	2	0.983	7	0.942	10
GR	0.45	12	0.899	9	0.977	7
IE	0.55	10	0.890	11	1.000	1
IT	0.41	13	0.672	14	0.700	14
LU	0.67	7	0.897	10	1.000	1
NL	0.83	2	1.000	1	1.000	1
PT	0.16	15	0.374	15	0.393	15
SE	0.91	1	1.000	1	1.000	1
UK	0.54	11	0.938	8	0.869	12
Mean	**0.62**		**0.893**		**0.871**	

Table 4: Sensitivity Analysis – Social Protection Index, 2006

	SPI1 Scores	rank	*SPI2* Scores	rank	*SPI3* Scores	rank
AT	0.82	4	1.22	3	0.79	1
BE	0.57	9	0.91	9	0.76	7
DE	0.58	8	0.90	11	0.76	7
DK	0.75	6	1.40	1	0.79	1
ES	0.40	14	0.68	14	0.71	14
FI	0.78	5	1.19	4	0.79	1
FR	0.83	2	1.07	6	0.78	5
GR	0.45	12	0.91	9	0.73	12
IE	0.55	10	0.89	12	0.75	10
IT	0.41	13	0.73	13	0.73	12
LU	0.67	7	1.03	7	0.76	7
NL	0.83	2	1.15	5	0.78	5
PT	0.16	15	0.65	15	0.68	15
SE	0.91	1	1.25	2	0.79	1
UK	0.54	11	1.02	8	0.75	10
Mean	**0.62**		**1.00**		**0.76**	

Note: SPI1, SPI2 and SPI3 results correspond to HDI, Afonso *et al.* and "goalpost" normalization data, respectively.

Table 5: Sensitivity Analysis – DEA Efficiency Scores, 2006

	*DEA*1 Scores	rank	*DEA*2 Scores	rank	*DEA*3 Scores	rank
AT	1.000	1	1.000	1	1.000	1
BE	0.866	13	0.981	11	0.978	14
DE	0.879	12	0.986	9	0.982	12
DK	1.000	1	1.000	1	1.000	1
ES	1.000	1	1.000	1	1.000	1
FI	1.000	1	1.000	1	1.000	1
FR	0.983	7	0.999	7	0.998	7
GR	0.899	9	0.981	11	0.995	9
IE	0.890	11	0.984	10	0.984	11
IT	0.672	14	0.988	8	0.980	13
LU	0.897	10	0.980	13	0.995	9
NL	1.000	1	1.000	1	1.000	1
PT	0.374	15	0.973	15	0.972	15
SE	1.000	1	1.000	1	1.000	1
UK	0.938	8	0.979	14	0.997	8
Mean	**0.893**		**0.990**		**0.992**	

Note: DEA1, DEA2 and DEA3 results correspond to HDI, Afonso *et al.* and "goalpost" normalization data, respectively.

Table 6: Sensitivity Analysis – DEA with Social Expenditures as Input, 2006

	DEA1-I Scores	rank	*DEA2-I* Scores	rank	*DEA3-I* Scores	rank
AT	0.972	8	0.940	7	0.845	7
BE	0.744	13	0.733	14	0.717	14
DE	0.872	11	0.824	11	0.776	11
DK	0.946	9	1.000	1	0.888	5
ES	1.000	1	0.917	8	0.909	4
FI	1.000	1	1.000	1	0.927	3
FR	0.942	10	0.756	13	0.707	15
GR	0.977	7	0.962	6	0.877	6
IE	1.000	1	1.000	1	1.000	1
IT	0.700	14	0.733	14	0.750	12
LU	1.000	1	1.000	1	1.000	1
NL	1.000	1	1.000	1	0.841	7
PT	0.393	15	0.777	12	0.783	10
SE	1.000	1	0.858	10	0.818	9
UK	0.869	12	0.881	9	0.725	13
Mean	**0.894**		**0.892**		**0.838**	

Note: DEA1-I, DEA2-I and DEA3-I results correspond to HDI, Afonso *et al.* and "goalpost" normalization data, respectively.

Table 7: Means of the DEA implicit weights

	POV	INE	UNE	EDU	EXP
DEA1	0.062	0.067	0.237	0.460	0.174
DEA2	0.080	0.080	0.072	0.030	0.738
DEA3	0.047	0.100	0.419	0.101	0.333

Table 8: Correlations between indexes

	SPI1	*SPI2*	*SPI3*	*DEA1*	*DEA2*	*DEA3*
SPI1	1.000					
SPI2	0.884	1				
SPI3	0.968	0.895	1.000			
DEA1	0.778	0.671	0.770	1.000		
DEA2	0.708	0.589	0.630	0.685	1.000	
DEA3	0.692	0.689	0.593	0.836	0.721	1.000

Table 9: Average indicator (*SPI1*) and DEA measures (*DEA1*) - 1995-2006

	DEA1											
	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
AT	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.990	1.000	1.000	1.000
BE	0.803	0.862	0.851	0.819	0.791	0.877	0.890	0.910	0.882	0.915	0.904	0.907
DE	0.839	0.852	0.846	0.902	0.874	1.000	0.927	0.912	0.879	0.904	0.906	0.918
DK	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.987	1.000	1.000	1.000
ES	0.770	0.764	0.829	0.833	0.813	0.895	0.862	0.950	1.000	0.965	0.925	1.000
FI	1.000	1.000	1.000	1.000	0.929	1.000	0.984	1.000	1.000	1.000	0.993	1.000
FR	0.808	0.803	0.829	0.833	0.860	0.868	0.917	0.920	0.906	0.966	0.970	0.997
GR	0.711	0.685	0.731	0.652	0.705	0.730	0.796	0.804	0.813	0.830	0.888	0.954
IE	0.626	0.709	0.696	0.753	0.855	0.916	0.937	0.937	0.935	0.949	0.962	0.925
IT	0.836	0.842	0.853	0.857	0.930	0.987	0.975	1.000	0.869	1.000	0.946	0.843
LU	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.993	0.952
NL	0.860	0.871	0.971	1.000	0.991	1.000	1.000	1.000	0.991	0.951	0.923	1.000
PT	0.747	0.737	0.755	0.862	0.885	0.887	0.907	0.899	0.866	0.781	0.711	0.701
SE	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
UK	0.738	0.803	0.874	0.913	0.907	0.925	0.928	0.960	0.981	1.000	1.000	0.971

	SPI1											
	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
AT	0.690	0.699	0.739	0.758	0.768	0.798	0.809	0.797	0.787	0.803	0.827	0.832
BE	0.528	0.578	0.614	0.605	0.629	0.664	0.692	0.678	0.683	0.689	0.678	0.688
DE	0.574	0.618	0.663	0.691	0.701	0.738	0.745	0.648	0.643	0.636	0.706	0.700
DK	0.748	0.731	0.759	0.777	0.779	0.789	0.809	0.791	0.752	0.793	0.803	0.805
ES	0.278	0.321	0.310	0.394	0.427	0.495	0.499	0.527	0.547	0.534	0.547	0.529
FI	0.719	0.730	0.758	0.765	0.747	0.768	0.763	0.776	0.795	0.761	0.799	0.814
FR	0.606	0.621	0.631	0.644	0.651	0.672	0.742	0.760	0.751	0.725	0.762	0.822
GR	0.375	0.397	0.399	0.379	0.395	0.434	0.458	0.438	0.417	0.486	0.518	0.565
IE	0.353	0.384	0.421	0.458	0.504	0.534	0.563	0.571	0.583	0.591	0.643	0.654
IT	0.342	0.364	0.409	0.453	0.483	0.519	0.519	0.528	0.529	0.570	0.564	0.550
LU	0.598	0.614	0.662	0.678	0.714	0.753	0.742	0.761	0.793	0.793	0.795	0.732
NL	0.670	0.652	0.750	0.769	0.754	0.751	0.771	0.774	0.761	0.772	0.796	0.840
PT	0.178	0.234	0.234	0.244	0.289	0.319	0.335	0.300	0.326	0.293	0.349	0.389
SE	0.873	0.874	0.880	0.875	0.906	0.861	0.894	0.872	0.889	0.892	0.915	0.877
UK	0.399	0.466	0.517	0.513	0.536	0.565	0.582	0.585	0.609	0.634	0.609	0.635

Table 10: Malmquist TFP indices

	1995-1996	1996-1997	1997-1998	1998-1999	1999-2000	2000-2001	2001-2002	2002-2003	2003-2004	2004-2005	2005-2006	Average
AT	-0.9%	0.6%	1.7%	1.2%	2.5%	1.2%	0.7%	0.2%	0.4%	-0.9%	-0.7%	0.60%
BE	6.8%	0.6%	-4.6%	-2.8%	7.7%	-2.9%	3.5%	-0.2%	3.7%	-2.9%	1.7%	0.90%
DE	0.4%	0.8%	2.4%	0.4%	4.9%	-3.2%	-4.6%	-0.9%	2.9%	-0.1%	-0.6%	0.20%
DK	-2.4%	2.6%	0.2%	0.6%	-0.1%	2.3%	-3.3%	-2.3%	0.7%	-0.7%	0.0%	-0.20%
ES	4.7%	17.1%	3.0%	0.0%	14.7%	1.5%	10.1%	17.1%	-1.6%	-0.8%	14.7%	7.10%
FI	-0.1%	0.0%	-3.4%	-5.8%	2.3%	-2.7%	0.7%	2.5%	-0.6%	-2.5%	1.3%	-0.80%
FR	1.8%	7.9%	2.4%	5.7%	5.3%	2.3%	0.5%	-0.2%	7.8%	2.6%	5.7%	3.80%
GR	-1.9%	11.1%	-9.2%	8.0%	1.4%	2.6%	2.6%	3.9%	2.0%	5.1%	8.9%	3.00%
IE	10.5%	0.0%	9.2%	15.3%	9.0%	3.3%	-0.4%	-1.0%	-0.2%	1.6%	-3.0%	3.90%
IT	6.3%	9.4%	2.9%	11.1%	10.6%	4.0%	2.6%	-4.7%	20.4%	-4.9%	-5.5%	4.50%
LU	0.4%	1.2%	0.0%	3.4%	1.8%	-0.2%	-0.9%	3.0%	-4.7%	-0.8%	-2.3%	0.00%
NL	-0.6%	11.2%	6.0%	1.5%	3.0%	2.1%	-0.8%	-2.6%	-5.8%	1.0%	3.9%	1.60%
PT	-2.4%	1.4%	14.1%	4.9%	1.3%	2.3%	-1.9%	-5.6%	-10.7%	-9.0%	-0.1%	-0.70%
SE	1.5%	2.8%	-1.4%	3.6%	-6.0%	3.4%	-6.4%	3.0%	0.1%	5.1%	-2.7%	0.20%
UK	7.5%	10.0%	7.3%	3.2%	4.1%	1.4%	2.4%	0.3%	1.1%	0.0%	-1.5%	3.20%

Figure 1: Average indicator SPI1 1995-2006

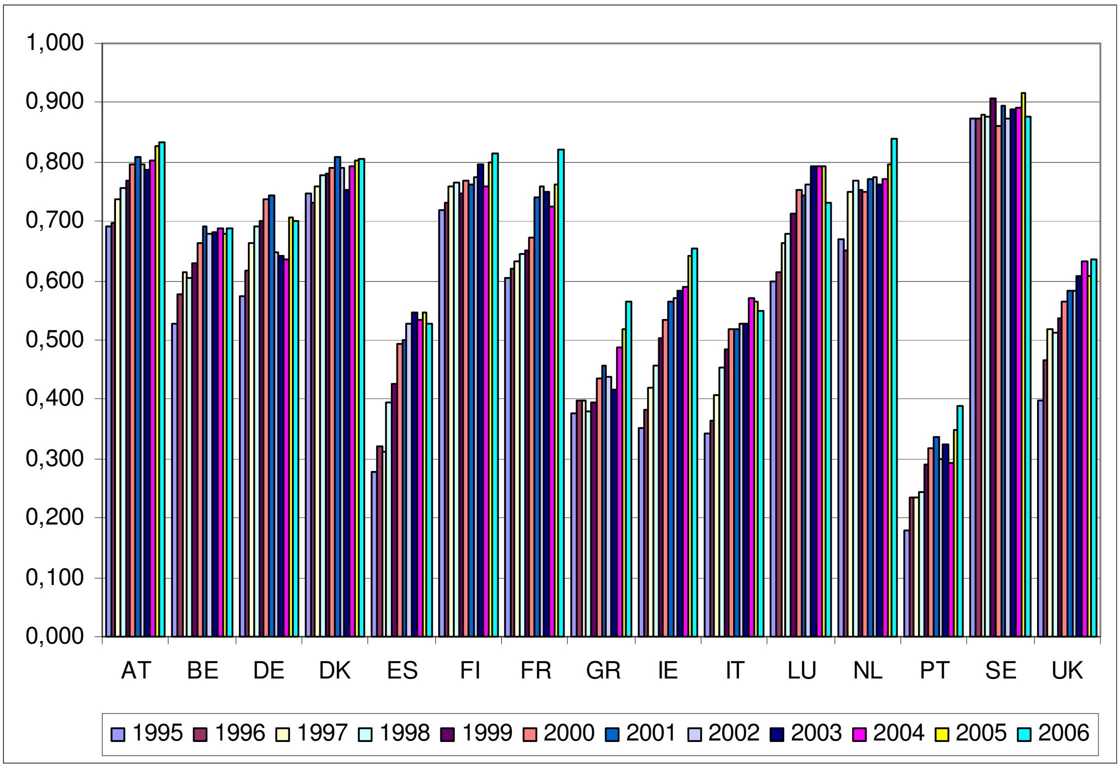


Figure 2: Convergence of SPI1

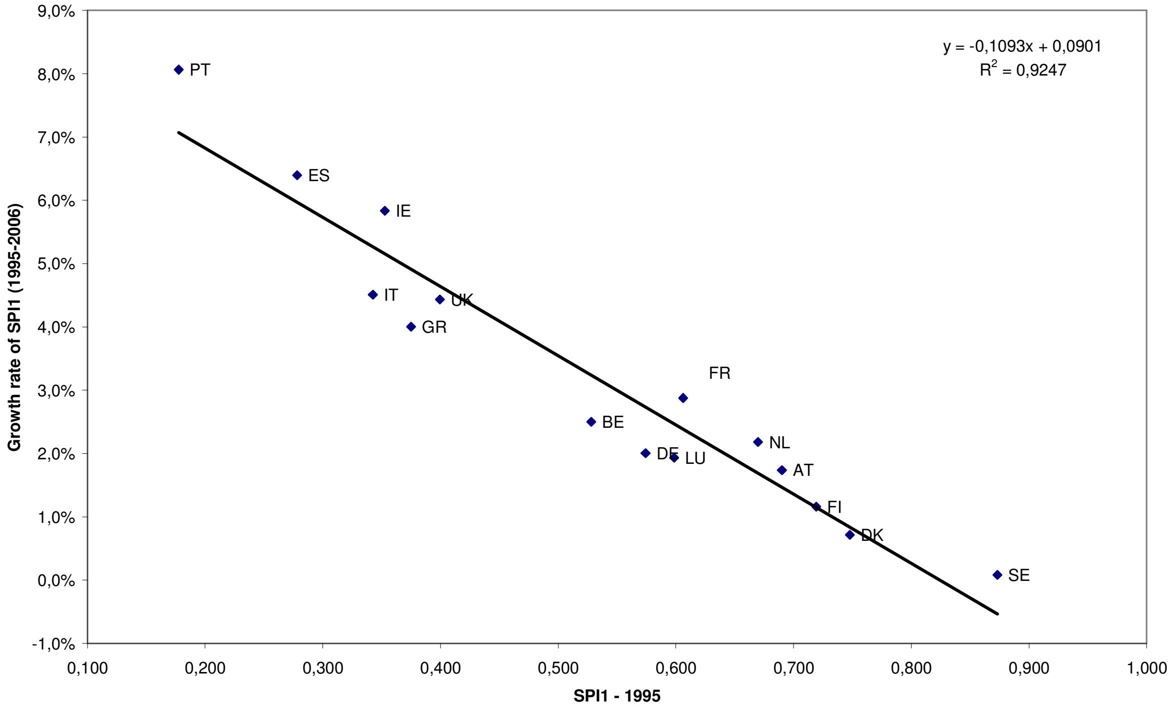


Figure 3: Convergence of *DEA1* according to Malmquist TFP change

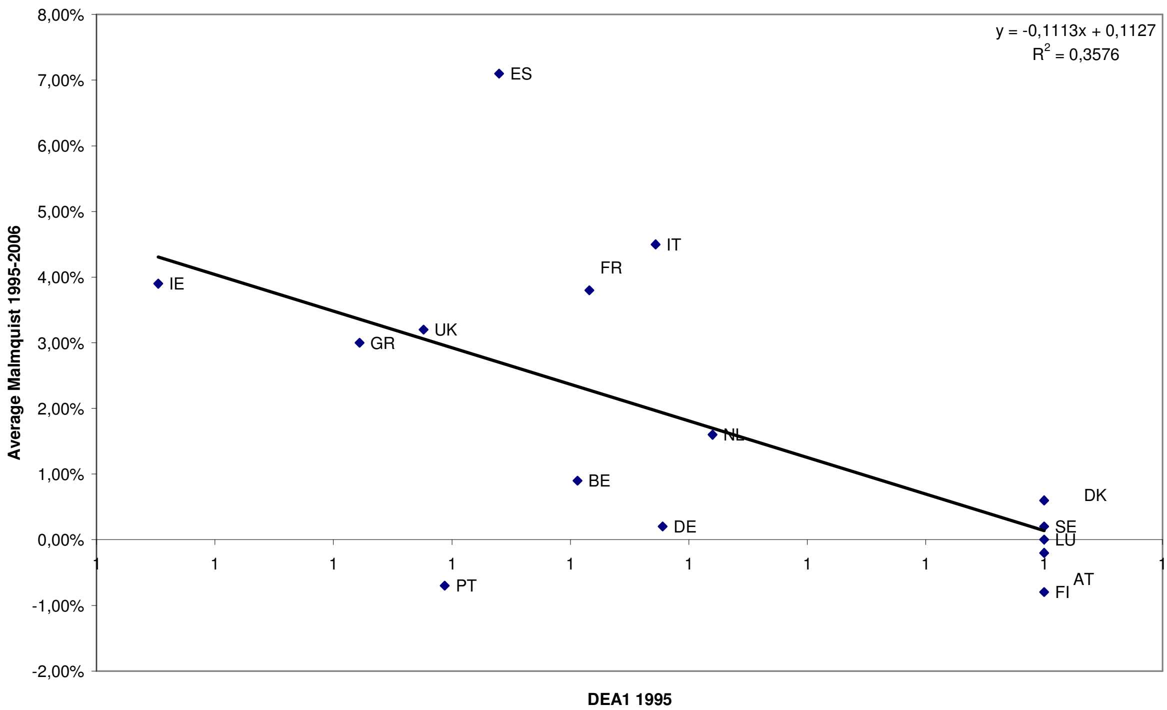
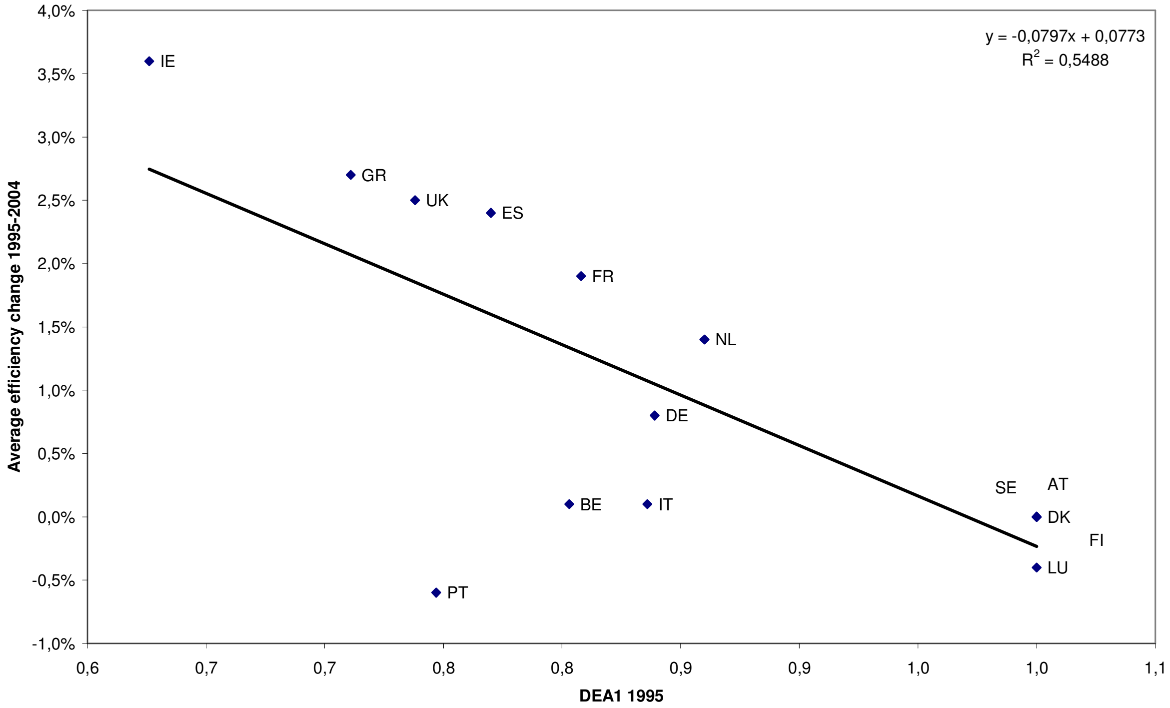


Figure 4: Convergence of *DEA1* according to "technical efficiency" change



Appendices

A.1. Social cohesion indicators

Table A1: Social cohesion indicators

	Primary indicators					HDR normalization				
	POV	INE	UNE	EDU	EXP	POV	INE	UNE	EDU	EXP
					1995					
AT	13	4.0	1.0	13.6	76.7	0.67	0.76	0.96	0.81	0.25
BE	16	4.5	5.8	15.1	76.9	0.47	0.64	0.46	0.78	0.29
DE	15	4.6	3.9	13.5	76.6	0.53	0.62	0.66	0.82	0.24
DK	10	2.9	2.0	6.1	75.3	0.87	1.00	0.86	1.00	0.02
ES	19	5.9	10.3	33.8	78.0	0.27	0.33	0.00	0.32	0.47
FI	8	3.0	5.3	11.5	76.6	1.00	0.98	0.52	0.87	0.24
FR	15	4.5	4.4	15.4	78.0	0.53	0.64	0.61	0.77	0.47
GR	22	6.5	4.6	22.4	77.7	0.07	0.20	0.59	0.60	0.42
IE	19	5.1	7.6	21.4	75.7	0.27	0.51	0.28	0.62	0.08
IT	20	5.9	7.1	32.8	78.2	0.20	0.33	0.33	0.34	0.51
LU	12	4.3	0.7	33.4	76.7	0.73	0.69	0.99	0.33	0.25
NL	11	4.2	3.1	18.0	77.5	0.80	0.71	0.74	0.71	0.39
PT	23	7.4	3.1	41.4	75.3	0.00	0.00	0.74	0.13	0.02
SE	8	3.0	2.3	8.0	78.8	1.00	0.98	0.82	0.95	0.61
UK	20	5.2	3.5	32.3	76.7	0.20	0.49	0.70	0.35	0.25
					1996					
AT	14	3.8	1.2	12.1	77.0	0.60	0.80	0.94	0.85	0.31
BE	15	4.2	5.7	12.9	77.2	0.53	0.71	0.47	0.83	0.34
DE	14	4.0	4.1	13.3	76.8	0.60	0.76	0.64	0.82	0.27
DK	10	3.0	1.8	12.1	75.7	0.87	0.98	0.88	0.85	0.08
ES	18	6.0	9.4	31.4	78.1	0.33	0.31	0.09	0.38	0.49
FI	8	3.0	5.2	11.1	76.8	1.00	0.98	0.53	0.88	0.27
FR	15	4.3	4.5	15.2	78.2	0.53	0.69	0.60	0.78	0.51
GR	21	6.3	5.2	20.7	77.8	0.13	0.24	0.53	0.64	0.44
IE	19	5.1	7.0	18.9	75.9	0.27	0.51	0.34	0.68	0.12
IT	20	5.6	7.3	31.7	78.4	0.20	0.40	0.31	0.37	0.54
LU	11	4.0	0.8	35.3	76.7	0.80	0.76	0.98	0.28	0.25
NL	12	4.4	3.0	17.6	77.5	0.73	0.67	0.75	0.72	0.39
PT	21	6.7	3.3	40.1	75.2	0.13	0.16	0.72	0.16	0.00
SE	8	3.0	2.7	7.5	79.0	1.00	0.98	0.78	0.97	0.64
UK	18	5.0	3.1	29.2	76.9	0.33	0.53	0.74	0.43	0.29
					1997					
AT	13	3.6	1.3	10.8	77.4	0.67	0.84	0.93	0.88	0.37
BE	14	4.0	5.4	12.7	77.4	0.60	0.76	0.51	0.84	0.37
DE	12	3.7	4.6	12.9	77.2	0.73	0.82	0.59	0.83	0.34
DK	10	2.9	1.5	10.7	76.0	0.87	1.00	0.91	0.89	0.14
ES	20	6.5	8.7	30.0	78.6	0.20	0.20	0.16	0.41	0.58
FI	8	3.0	4.9	8.1	77.0	1.00	0.98	0.56	0.95	0.31
FR	15	4.4	4.7	14.1	78.6	0.53	0.67	0.58	0.80	0.58
GR	21	6.6	5.3	19.9	78.2	0.13	0.18	0.52	0.66	0.51
IE	19	5.0	5.6	18.9	76.0	0.27	0.53	0.48	0.68	0.14
IT	19	5.3	7.3	30.1	78.7	0.27	0.47	0.31	0.41	0.59
LU	11	3.6	0.9	30.7	77.0	0.80	0.84	0.97	0.39	0.31
NL	10	3.6	2.3	16.0	77.9	0.87	0.84	0.82	0.76	0.46
PT	22	6.7	3.2	40.6	75.6	0.07	0.16	0.73	0.15	0.07
SE	8	3.0	3.1	6.8	79.3	1.00	0.98	0.74	0.98	0.69
UK	18	4.7	2.5	26.0	77.2	0.33	0.60	0.80	0.51	0.34
					1998					
AT	13	3.5	1.3	10.7	77.8	0.67	0.87	0.93	0.89	0.44
BE	14	4.0	5.6	14.5	77.5	0.60	0.76	0.48	0.79	0.39
DE	11	3.6	4.5	13.9	77.6	0.80	0.84	0.60	0.81	0.41
DK	10	3.0	1.3	9.8	76.4	0.87	0.98	0.93	0.91	0.20
ES	18	5.9	7.5	29.6	78.7	0.33	0.33	0.29	0.42	0.59
FI	9	3.1	4.1	7.9	77.2	0.93	0.96	0.64	0.96	0.34

	Primary indicators					HDR normalization				
	POV	INE	UNE	EDU	EXP	POV	INE	UNE	EDU	EXP
FR	15	4.2	4.5	14.9	78.7	0.53	0.71	0.60	0.78	0.59
GR	21	6.5	5.8	20.7	77.9	0.13	0.20	0.46	0.64	0.46
IE	19	5.2	3.9	18.0	76.2	0.27	0.49	0.66	0.71	0.17
IT	18	5.1	6.8	28.4	78.8	0.33	0.51	0.36	0.45	0.61
LU	12	3.7	0.9	25.2	77.2	0.73	0.82	0.97	0.53	0.34
NL	10	3.6	1.5	15.5	77.9	0.87	0.84	0.91	0.77	0.46
PT	21	6.8	2.2	46.6	75.9	0.13	0.13	0.84	0.00	0.12
SE	8	3.4	2.6	6.9	79.4	1.00	0.89	0.79	0.98	0.71
UK	19	5.2	1.9	22.9	77.3	0.27	0.49	0.87	0.59	0.36
1999										
AT	12	3.7	1.2	10.7	77.9	0.73	0.82	0.94	0.89	0.46
BE	13	4.2	4.8	15.2	77.7	0.67	0.71	0.57	0.78	0.42
DE	11	3.6	4.1	14.9	77.8	0.80	0.84	0.64	0.78	0.44
DK	10	3.0	1.1	11.5	76.6	0.87	0.98	0.95	0.87	0.24
ES	19	5.7	5.7	29.5	78.7	0.27	0.38	0.47	0.42	0.59
FI	11	3.4	3.0	9.9	77.5	0.80	0.89	0.75	0.91	0.39
FR	15	4.4	4.1	14.7	78.9	0.53	0.67	0.64	0.79	0.63
GR	21	6.2	6.5	18.6	78.1	0.13	0.27	0.39	0.69	0.49
IE	19	4.9	2.4	17.1	76.1	0.27	0.56	0.81	0.73	0.15
IT	18	4.9	6.7	27.2	79.2	0.33	0.56	0.37	0.48	0.68
LU	13	3.9	0.7	19.1	77.9	0.67	0.78	0.99	0.68	0.46
NL	11	3.7	1.2	16.2	77.9	0.80	0.82	0.94	0.75	0.46
PT	21	6.4	1.8	44.9	76.2	0.13	0.22	0.88	0.04	0.17
SE	8	3.1	1.9	6.9	79.5	1.00	0.96	0.87	0.98	0.73
UK	19	5.2	1.7	19.7	77.4	0.27	0.49	0.89	0.66	0.37
2000										
AT	12	3.4	1.0	10.2	78.2	0.73	0.89	0.96	0.90	0.51
BE	13	4.3	3.7	12.5	77.8	0.67	0.69	0.68	0.84	0.44
DE	10	3.5	3.7	14.9	78.1	0.87	0.87	0.68	0.78	0.49
DK	10	3.1	0.9	11.6	76.9	0.87	0.96	0.97	0.86	0.29
ES	18	5.4	4.6	29.1	79.2	0.33	0.44	0.59	0.43	0.68
FI	11	3.3	2.8	8.9	77.7	0.80	0.91	0.77	0.93	0.42
FR	16	4.2	3.5	13.3	79.1	0.47	0.71	0.70	0.82	0.66
GR	20	5.8	6.2	18.2	78.1	0.20	0.36	0.42	0.70	0.49
IE	20	4.7	1.6	16.2	76.5	0.20	0.60	0.90	0.75	0.22
IT	18	4.8	6.3	25.3	79.6	0.33	0.58	0.41	0.53	0.75
LU	12	3.7	0.6	16.8	78.0	0.73	0.82	1.00	0.74	0.47
NL	11	4.1	0.8	15.5	78.0	0.80	0.73	0.98	0.77	0.47
PT	21	6.4	1.7	42.6	76.7	0.13	0.22	0.89	0.10	0.25
SE	11	3.5	1.4	7.7	79.7	0.80	0.87	0.92	0.96	0.76
UK	19	5.2	1.4	18.4	77.9	0.27	0.49	0.92	0.70	0.46
2001										
AT	12	3.5	0.9	10.2	78.6	0.73	0.87	0.97	0.90	0.58
BE	13	4	3.2	13.6	78.1	0.67	0.76	0.73	0.81	0.49
DE	11	3.6	3.7	12.5	78.5	0.80	0.84	0.68	0.84	0.56
DK	10	3	0.9	9.0	77.0	0.87	0.98	0.97	0.93	0.31
ES	19	5.5	3.7	29.2	79.3	0.27	0.42	0.68	0.43	0.69
FI	11	3.7	2.5	10.3	78.1	0.80	0.82	0.80	0.90	0.49
FR	13	3.9	3.0	13.5	79.3	0.67	0.78	0.75	0.82	0.69
GR	20	5.7	5.5	17.3	78.1	0.20	0.38	0.49	0.72	0.49
IE	21	4.5	1.3	15.3	77.2	0.13	0.64	0.93	0.77	0.34
IT	19	4.8	5.7	26.4	79.8	0.27	0.58	0.47	0.50	0.78
LU	12	3.8	0.6	18.1	78.0	0.73	0.80	1.00	0.70	0.47
NL	11	4	0.6	15.3	78.3	0.80	0.76	1.00	0.77	0.53
PT	20	6.5	1.5	44.0	77.0	0.20	0.20	0.91	0.06	0.31
SE	9	3.4	1.0	10.5	79.9	0.93	0.89	0.96	0.89	0.80
UK	18	5.4	1.3	17.7	78.1	0.33	0.44	0.93	0.71	0.49
2002										
AT	12.5	3.75	1.1	9.5	78.8	0.70	0.81	0.95	0.92	0.61
BE	14	4	3.7	12.4	78.2	0.60	0.76	0.68	0.84	0.51
DE	15	4.4	3.9	12.6	78.4	0.53	0.67	0.66	0.84	0.54
DK	11	3.3	0.9	8.6	77.2	0.80	0.91	0.97	0.94	0.34
ES	19	5.1	3.7	29.9	79.7	0.27	0.51	0.68	0.41	0.76
FI	11	3.7	2.3	9.9	78.3	0.80	0.82	0.82	0.91	0.53

	Primary indicators					HDR normalization				
	POV	INE	UNE	EDU	EXP	POV	INE	UNE	EDU	EXP
FR	12	3.9	3.1	13.4	79.5	0.73	0.78	0.74	0.82	0.73
GR	20.5	6.15	5.3	16.7	78.1	0.17	0.28	0.52	0.74	0.49
IE	21	4.8	1.4	14.7	77.8	0.13	0.58	0.92	0.79	0.44
IT	19	5.2	5.1	24.3	79.9	0.27	0.49	0.54	0.55	0.80
LU	11	3.9	0.7	17	78.2	0.80	0.78	0.99	0.73	0.51
NL	11	4	0.7	15	78.4	0.80	0.76	0.99	0.78	0.54
PT	20	7.3	1.7	45.1	77.3	0.20	0.02	0.89	0.04	0.36
SE	11	3.3	1	10.4	79.9	0.80	0.91	0.96	0.89	0.80
UK	18	5.5	1.1	17.8	78.2	0.33	0.42	0.95	0.71	0.51
2003										
AT	13	4	1.1	9.3	79.0	0.67	0.76	0.95	0.92	0.64
BE	15	4	3.7	12.8	78.8	0.53	0.76	0.68	0.83	0.61
DE	15	4.3	4.5	12.8	78.5	0.53	0.69	0.60	0.83	0.56
DK	12	3.6	1.1	10.3	77.2	0.73	0.84	0.95	0.90	0.34
ES	19	5.1	3.7	31.3	80.5	0.27	0.51	0.68	0.38	0.90
FI	11	3.6	2.3	8.3	78.5	0.80	0.84	0.82	0.95	0.56
FR	12	3.8	3.7	13.7	79.5	0.73	0.80	0.68	0.81	0.73
GR	21	6.6	5.3	15.5	78.1	0.13	0.18	0.52	0.77	0.49
IE	21	5.1	1.6	12.3	78.3	0.13	0.51	0.90	0.85	0.53
IT	19	5.2	4.9	23.5	79.7	0.27	0.49	0.56	0.57	0.76
LU	10	4	0.9	12.3	78.3	0.87	0.76	0.97	0.85	0.53
NL	12	4	1	14.2	78.5	0.73	0.76	0.96	0.80	0.56
PT	19	7.4	2.2	40.4	77.4	0.27	0.00	0.84	0.15	0.37
SE	11	3.3	1	9	80.2	0.80	0.91	0.96	0.93	0.85
UK	18	5.3	1.1	16.8	78.5	0.33	0.47	0.95	0.74	0.56
2004										
AT	13	3.8	1.3	8.7	79.3	0.67	0.80	0.93	0.94	0.69
BE	15	4	4.1	11.9	79.1	0.53	0.76	0.64	0.86	0.66
DE	16	4.4	5.4	12.1	79.3	0.47	0.67	0.51	0.85	0.69
DK	11	3.4	1.2	8.5	77.6	0.80	0.89	0.94	0.94	0.40
ES	20	5.1	3.4	31.7	80.4	0.20	0.51	0.71	0.37	0.88
FI	11	3.5	2.1	8.7	77.3	0.80	0.87	0.85	0.94	0.36
FR	14	4.2	3.9	14.2	80.3	0.60	0.71	0.66	0.80	0.86
GR	20	6	5.6	14.9	79.1	0.20	0.31	0.48	0.78	0.65
IE	21	5	1.6	12.9	78.5	0.13	0.53	0.90	0.83	0.56
IT	19	5.6	4	22.3	80.7	0.27	0.40	0.65	0.60	0.93
LU	11	3.7	1.1	12.7	78.5	0.80	0.82	0.95	0.84	0.56
NL	12	4	1.6	14	79.2	0.73	0.76	0.90	0.80	0.67
PT	21	7.2	3	39.4	77.3	0.13	0.04	0.75	0.18	0.36
SE	11	3.3	1.2	8.6	80.4	0.80	0.91	0.94	0.94	0.87
UK	18	5.3	1	14.9	78.9	0.33	0.47	0.96	0.78	0.63
2005										
AT	12	3.8	1.3	9	79.6	0.73	0.80	0.93	0.93	0.75
BE	15	4	4.4	13	79.1	0.53	0.76	0.61	0.83	0.66
DE	12	3.8	5.7	13.8	79.4	0.73	0.80	0.47	0.81	0.71
DK	12	3.5	1.1	8.5	78.3	0.73	0.87	0.95	0.94	0.53
ES	20	5.4	2.2	30.8	80.3	0.20	0.44	0.84	0.39	0.86
FI	12	3.6	2.2	9.3	79.1	0.73	0.84	0.84	0.92	0.66
FR	13	4	3.8	12	80.3	0.67	0.76	0.67	0.85	0.86
GR	20	5.8	5.1	13.3	79.2	0.20	0.36	0.54	0.82	0.68
IE	20	5	1.5	12.3	79.5	0.20	0.53	0.91	0.85	0.73
IT	19	5.6	3.9	21.9	80.4	0.27	0.40	0.66	0.61	0.88
LU	13	3.8	1.2	13.3	79.6	0.67	0.80	0.94	0.82	0.75
NL	11	4	1.9	13.6	79.6	0.80	0.76	0.87	0.81	0.75
PT	19	6.9	3.7	38.6	78.1	0.27	0.11	0.68	0.20	0.49
SE	9	3.3	1.2	11.7	80.7	0.93	0.91	0.94	0.86	0.93
UK	19	5.8	1	14	79.1	0.27	0.36	0.96	0.80	0.66
2006										
AT	13	3.7	1.3	9.6	80.1	0.67	0.82	0.93	0.91	0.83
BE	15	4.2	4.2	12.6	79.5	0.53	0.71	0.63	0.84	0.73
DE	13	4.1	5.5	13.9	79.9	0.67	0.73	0.49	0.81	0.80
DK	12	3.4	0.8	10.9	78.4	0.73	0.89	0.98	0.88	0.54
ES	20	5.3	4.8	29.9	81.1	0.20	0.47	0.57	0.41	1.00
FI	13	3.6	1.9	8.3	79.6	0.67	0.84	0.87	0.95	0.75

	Primary indicators					HDR normalization				
	POV	INE	UNE	EDU	EXP	POV	INE	UNE	EDU	EXP
FR	13	4	1.8	12.3	80.9	0.67	0.76	0.88	0.85	0.97
GR	21	6.1	1.4	15.9	79.5	0.13	0.29	0.92	0.76	0.73
IE	18	4.9	2.8	12.3	79.7	0.33	0.56	0.77	0.85	0.76
IT	20	5.5	3.9	20.8	80.1	0.20	0.42	0.66	0.64	0.83
LU	14	4.2	1.4	17.4	79.4	0.60	0.71	0.92	0.72	0.71
NL	10	3.8	1.7	12.9	80.0	0.87	0.80	0.89	0.83	0.81
PT	18	6.8	3.8	39.2	78.9	0.33	0.13	0.67	0.18	0.63
SE	12	3.5	1.1	12	81.0	0.73	0.87	0.95	0.85	0.98
UK	19	5.4	1.2	13	79.3	0.27	0.44	0.94	0.83	0.70

Source: Eurostat Laeken Indicators. Income and Living Conditions Database (2007).

A.2. Data envelopment analysis

In this appendix we describe the data envelopment analysis (DEA) methods that have been used in this paper.

Given access to data on N inputs and M outputs for each of I countries, a DEA model may be specified as[24]

$$
\begin{aligned}
&\max_{\lambda} \phi \\
&\text{st} \quad -\phi\mathbf{q}_i + \mathbf{Q\lambda} \geq \mathbf{0}, \\
&\qquad \mathbf{x}_i - \mathbf{X\lambda} \geq \mathbf{0}, \\
&\qquad \boldsymbol{\lambda} \geq \mathbf{0},
\end{aligned}
\tag{A1}
$$

where $\mathbf{x}_i$ is the input vector of the i-th country; $\mathbf{q}_i$, is the output vector of the i-th country; the $N\times I$ input matrix, $\mathbf{X}$, and the $M\times I$ output matrix, $\mathbf{Q}$, represent the data for all I countrys; ϕ is a scalar and $\boldsymbol{\lambda}$ is a $I\times1$ vector of constants. The value of ϕ obtained is the inverse of the efficiency score for the i-th country. It satisfies: $1\leq\phi\leq\infty$, with a value of 1 indicating a point on the frontier and hence a technically efficient country. Note that the linear programming problem is solved I times, once for each country in the sample. A value of ϕ is then obtained for each country.

Note that, in the event that all countries have a single unit of input, which is the case in our situation, the LP in (A1) reduces to

$$
\begin{aligned}
&\max_{\lambda} \phi \\
&\text{st} \quad -\phi\mathbf{q}_i + \mathbf{Q\lambda} \geq \mathbf{0},
\end{aligned}
$$

[24] This is an output oriented constant returns to scale DEA model. See, for example, Färe *et al.* (1985).

$$\boldsymbol{\lambda} \geq \mathbf{0}, \tag{A2}$$

Shadow prices

Then the shadow price information that is produced by a DEA model can be illustrated by considering the dual to the output-oriented DEA LP problem in (A1)[25]

$$\begin{aligned} &\min_{\boldsymbol{\mu},\boldsymbol{\nu}} (\boldsymbol{\nu}'\mathbf{x}_i), \\ \text{st} \quad &\boldsymbol{\mu}'\mathbf{q}_i = 1, \\ &\boldsymbol{\mu}'\mathbf{q}_j - \boldsymbol{\nu}'\mathbf{x}_j \leq 0, \qquad j=1,2,...,I, \\ &\boldsymbol{\mu}, \boldsymbol{\nu} \geq \mathbf{0}, \end{aligned} \tag{A3}$$

where $\boldsymbol{\mu}$ is an $M\times1$ vector of output shadow prices and $\boldsymbol{\nu}$ is a $N\times1$ vector of input shadow prices, which correspond to the M output constraints and N input constraints in the primal LP in (7). [26] Once again, this LP is solved I times (once for each country in the sample) and the technical efficiency score of the i-th country will be equal to $\boldsymbol{\mu}'\mathbf{q}_i/\boldsymbol{\nu}'\mathbf{x}_i$, which will be identical to the ϕ obtained using the primal DEA LP (a standard duality result in linear programming).[27]

If one now considers the case where each country has one unit of a single input, we see that the efficiency score becomes $\boldsymbol{\mu}'\mathbf{q}_i/\nu_1$, which is a simple linear function of the q_i. The elements of $\boldsymbol{\mu}$ may be interpreted as normalized shadow prices. Thus the ratio of any two elements of $\boldsymbol{\mu}$ can be interpreted in the same way as equation (6) in the main text.

[25] The seminal DEA paper by Charnes, Cooper and Rhodes (1978) used an input-oriented dual formulation.
[26] See Coelli et al (2005, ch6) for discussion of primal and dual DEA LPs.
[27] Note that there is no need to solve both the primal and dual LPs. The shadow prices can be obtained directly from the final solution matrix when the primal LP is solved.

With the DEA method, the weights can vary from country to country when the output mix varies. For example, consider Figure A.1 where we illustrate a simple case where there are six countries with two output indicators. Countries A, B and C define the frontier and hence are fully efficient, while countries D, E and F are inside the frontier and hence inefficient. Country F has a technical efficiency score of 0F/0F*=0.7, indication that it is producing 70% of its potential output. The slope of the frontier is equal to $-\mu_2/\mu_1$. The slope of the line AB is 1 while that of BC is 2. Thus we could say that country F (and country E) allocates weights of 0.33 and 0.67 to outputs 1 and 2, respectively, while country D allocates equal weights of 0.5 to the two outputs.[28,29]

The Malmquist TFP Index

Following Färe *et al.* (1994), the Malmquist TFP index measures the TFP change between two data points (in periods *t* and *s*) by calculating the ratio of the distances of each data point relative to a common technology. If the period *t* technology is used as the reference technology, the Malmquist (output-orientated) TFP change index between period *s* (the base period) and period t is can be written as

$$m_o^t\left(\mathbf{q}_s,\mathbf{x}_s,\mathbf{q}_t,\mathbf{x}_t\right)=\frac{d_o^t\left(\mathbf{q}_t,\mathbf{x}_t\right)}{d_o^t\left(\mathbf{q}_s,\mathbf{x}_s\right)}. \tag{A4}$$

[28] We have scaled the weights so that they add up to one to make the discussion more easily comparable to the index numbers above.

[29] Note that the shadow prices may not be unique for technically efficient observations and for those inefficient observations that happen to project onto part of the frontier where two or more hyper-planes intersect. In this case our DEA algorithm simply reports the shadow prices from the final simplex tableau (corresponding to one of the hyper-planes that pass through that point). A better solution would involve the use of information from all the intersecting hyper-planes, however we have not yet developed an algorithm that can be used to derive a unique and defendable set of shadow prices from points such as these.

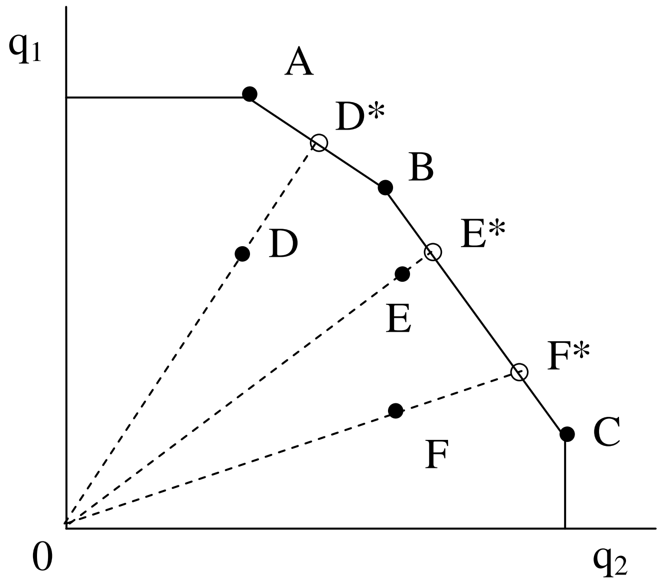


Figure A.1: DEA frontier

Alternatively, if the period s reference technology is used it is defined as

$$m_o^s\left(\mathbf{q}_s,\mathbf{x}_s,\mathbf{q}_t,\mathbf{x}_t\right)=\frac{d_o^s\left(\mathbf{q}_t,\mathbf{x}_t\right)}{d_o^s\left(\mathbf{q}_s,\mathbf{x}_s\right)}. \tag{A5}$$

Note that in the above equations the notation $d_o^s\left(\mathbf{q}_t,\mathbf{x}_t\right)$ represents the distance from the period *t* observation to the period s technology. A value of m_O greater than one indicates positive TFP growth from period *s* to period *t* while a value less than one indicates a TFP decline.

These two (period s and period t) indices are only equivalent if the technology is Hicks output neutral. To avoid the necessity to either impose this restriction or to arbitrarily choose one of the two technologies, the Malmquist TFP index is often defined as the geometric mean of these two indices

$$m_o(\mathbf{q}_s, \mathbf{x}_s, \mathbf{q}_t, \mathbf{x}_t) = \left[\frac{d_o^s(\mathbf{q}_t, \mathbf{x}_t)}{d_o^s(\mathbf{q}_s, \mathbf{x}_s)} \times \frac{d_o^t(\mathbf{q}_t, \mathbf{x}_t)}{d_o^t(\mathbf{q}_s, \mathbf{x}_s)} \right]^{1/2} \tag{A6}$$

The distance functions in this productivity index can be rearranged to show that it equivalent to the product of a technical efficiency change index and an index of technical change:

$$m_o(\mathbf{q}_s, \mathbf{x}_s, \mathbf{q}_t, \mathbf{x}_t) = \frac{d_o^t(\mathbf{q}_t, \mathbf{x}_t)}{d_o^s(\mathbf{q}_s, \mathbf{x}_s)} \left[\frac{d_o^s(\mathbf{q}_t, \mathbf{x}_t)}{d_o^t(\mathbf{q}_t, \mathbf{x}_t)} \times \frac{d_o^s(\mathbf{q}_s, \mathbf{x}_s)}{d_o^t(\mathbf{q}_s, \mathbf{x}_s)} \right]^{1/2} . \tag{A7}$$

The ratio outside the square brackets in the above equation measures the change in the output-oriented measure of Farrell technical efficiency between periods s and t. That is, the efficiency change index is equivalent to the ratio of the Farrell technical efficiency in period t to the Farrell technical efficiency in period s. The remaining part of the index in equation (A7) is a measure of technical change. It is the geometric mean of the shift in technology between the two periods, evaluated at $\mathbf{x}_t$ and also at $\mathbf{x}_s$.

Given that suitable panel data are available, we can calculate the four distance measures in equation (A7) using DEA-like linear programs. For the i-th country, we must calculate four distance functions to measure the TFP change between two periods. This requires the solving of four linear programming (LP) problems:[30]

$[d_o^t(\mathbf{q}_t, \mathbf{x}_t)]^{-1} = \max_{\phi,\boldsymbol{\lambda}} \phi,$

st $\quad -\phi\mathbf{q}_{it} + \mathbf{Q}_t\boldsymbol{\lambda} \geq \mathbf{0},$

[30] All notation follows directly from that used earlier. The only differences are that we now have time subscripts, s and t, to represent the two time periods of interest.

$$\mathbf{x}_{it} - \mathbf{X}_t\boldsymbol{\lambda} \geq \mathbf{0},$$

$$\boldsymbol{\lambda} \geq \mathbf{0}, \tag{A7}$$

$$[d_o^{\,s}(\mathbf{q}_s, \mathbf{x}_s)]^{-1} = \max_{\phi,\boldsymbol{\lambda}} \phi,$$

$$\text{st} \quad -\phi\mathbf{q}_{is} + \mathbf{Q}_s\boldsymbol{\lambda} \geq \mathbf{0},$$

$$\mathbf{x}_{is} - \mathbf{X}_s\boldsymbol{\lambda} \geq \mathbf{0},$$

$$\boldsymbol{\lambda} \geq \mathbf{0}, \tag{A8}$$

$$[d_o^{\,t}(\mathbf{q}_s, \mathbf{x}_t)]^{-1} = \max_{\phi,\boldsymbol{\lambda}} \phi,$$

$$\text{st} \quad -\phi\mathbf{q}_{is} + \mathbf{Q}_t\boldsymbol{\lambda} \geq \mathbf{0},$$

$$\mathbf{x}_{is} - \mathbf{X}_t\boldsymbol{\lambda} \geq \mathbf{0},$$

$$\boldsymbol{\lambda} \geq \mathbf{0}, \tag{A9}$$

and

$$[d_o^{\,s}(\mathbf{q}_t, \mathbf{x}_t)]^{-1} = \max_{\phi,\boldsymbol{\lambda}} \phi,$$

$$\text{st} \quad -\phi\mathbf{q}_{it} + \mathbf{Q}_s\boldsymbol{\lambda} \geq \mathbf{0},$$

$$\mathbf{x}_{it} - \mathbf{X}_s\boldsymbol{\lambda} \geq \mathbf{0},$$

$$\boldsymbol{\lambda} \geq \mathbf{0}. \tag{A10}$$

A.3. Social expenditures

Table A2: Social spending

	Social spending as a % of GDP											
Country	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
AT	28.7	28.6	28.6	28.3	28.7	28.2	28.6	29.1	29.5	29.1	28.8	28.5
BE	27.4	28	27.4	27.1	27	26.5	27.3	28	29.1	29.3	29.7	30.1
DE	28.2	29.3	28.9	28.8	29.2	29.2	29.3	29.9	30.2	30.7	29.4	28.2
DK	31.9	31.2	30.1	30	29.8	28.9	29.2	29.7	30.7	30.7	30.1	29.5
ES	21.6	21.5	20.8	20.2	19.8	19.7	19.5	19.8	19.9	20	20.8	21.6
FI	31.5	31.4	29.1	27	26.2	25.1	24.9	25.6	26.5	26.7	26.7	26.7
FR	30.3	30.6	30.4	30	29.9	29.5	29.6	30.4	30.9	31.2	31.4	31.6
GR	22.3	22.9	23.3	24.2	25.5	25.7	26.7	26.2	26	26	24.2	22.5
IE	18.8	17.6	16.4	15.2	14.6	14.1	15	16	16.5	17	18.2	19.5
IT	24.2	24.3	24.9	24.6	24.8	24.7	24.9	25.3	25.8	26.1	26.4	26.7
LU	20.7	21.2	21.5	21.2	20.5	19.6	20.8	21.4	22.2	22.6	21.9	21.2
NL	30.6	29.6	28.7	27.8	27.1	26.4	26.5	27.6	28.3	28.5	28.2	27.9
PT	21	20.2	20.3	20.9	21.4	21.7	22.7	23.7	24.2	24.9	24.9	24.9
SE	34.3	33.6	32.7	32	31.7	30.7	31.3	32.3	33.3	32.9	32	31.1
UK	28.2	28	27.5	26.9	26.4	27.1	27.5	26.4	26.4	26.3	26.8	27.3

Source: Eurostat (2007).

Table A3: Malmquist decomposition

	1995-1996		*1996-1997*		*1997-1998*		*1998-1999*		*1999-2000*		*2000-2001*		*2001-2002*		*2002-2003*		*2003-2004*		*2004-2005*		*2005-2006*	
	Eff. change	Tech. change	Eff. change	Tech. change	Eff. change	Tech. change	Eff. change	Tech. change	Eff. change	Tech. change	Eff. change	Tech. change	Eff. change	Tech. change	Eff. change	Tech. change	Eff. change	Tech. change	Eff. change	Tech. change	Eff. change	Tech. change
AT	0,0%	-1,0%	0,0%	0,0%	0,0%	1,0%	0,0%	1,0%	0,0%	2,0%	0,0%	1,0%	0,0%	0,0%	-1,0%	1,0%	1,1%	-1,0%	0,0%	-1,0%	0,0%	-1,0%
BE	7,3%	-1,0%	-1,2%	1,0%	-3,8%	-1,0%	-3,4%	0,0%	10,9%	-3,0%	1,4%	-5,0%	2,2%	1,0%	-3,0%	2,0%	3,7%	0,0%	-1,2%	-2,0%	0,3%	1,0%
DE	1,5%	-2,0%	-0,6%	1,0%	6,6%	-4,0%	-3,1%	3,0%	14,5%	-9,0%	-7,3%	4,0%	-1,7%	-3,0%	-3,6%	2,0%	2,9%	0,0%	0,2%	-1,0%	1,3%	-2,0%
DK	0,0%	-3,0%	0,0%	2,0%	0,0%	0,0%	0,0%	0,0%	0,0%	-1,0%	0,0%	2,0%	0,0%	-4,0%	-1,3%	-1,0%	1,3%	-1,0%	0,0%	-1,0%	0,0%	0,0%
ES	-0,8%	5,0%	8,5%	7,0%	0,5%	2,0%	-2,3%	2,0%	10,0%	4,0%	-3,6%	5,0%	10,1%	0,0%	5,3%	11,0%	-3,5%	2,0%	-4,1%	3,0%	8,1%	6,0%
FI	0,0%	-1,0%	0,0%	0,0%	0,0%	-4,0%	-7,1%	1,0%	7,6%	-5,0%	-1,6%	-2,0%	1,6%	-1,0%	0,0%	2,0%	0,0%	-1,0%	-0,7%	-2,0%	0,7%	0,0%
FR	-0,7%	2,0%	3,2%	4,0%	0,5%	1,0%	3,3%	2,0%	1,0%	4,0%	5,6%	-4,0%	0,3%	0,0%	-1,6%	1,0%	6,7%	1,0%	0,3%	2,0%	2,8%	2,0%
GR	-3,7%	1,0%	6,7%	4,0%	-10,7%	1,0%	8,1%	-1,0%	3,6%	-2,0%	8,9%	-6,0%	1,1%	1,0%	1,1%	2,0%	2,1%	-1,0%	7,0%	-2,0%	7,4%	1,0%
IE	13,3%	-3,0%	-1,8%	1,0%	8,3%	0,0%	13,5%	1,0%	7,0%	1,0%	2,4%	0,0%	0,0%	-1,0%	-0,2%	-1,0%	1,5%	-2,0%	1,4%	0,0%	-3,9%	1,0%
IT	0,7%	5,0%	1,4%	7,0%	0,4%	2,0%	8,6%	2,0%	6,1%	4,0%	-1,2%	5,0%	2,6%	0,0%	-13,1%	9,0%	15,1%	4,0%	-5,4%	0,0%	-10,9%	6,0%
LU	0,0%	0,0%	0,0%	1,0%	0,0%	0,0%	0,0%	3,0%	0,0%	1,0%	0,0%	-1,0%	0,0%	-1,0%	0,0%	3,0%	0,0%	-5,0%	-0,7%	-1,0%	-4,1%	1,0%
NL	1,2%	-2,0%	11,5%	-1,0%	3,0%	2,0%	-0,9%	2,0%	1,0%	2,0%	0,0%	2,0%	0,0%	-1,0%	-0,9%	-2,0%	-4,0%	-2,0%	-3,0%	4,0%	8,4%	-5,0%
PT	-1,3%	-2,0%	2,4%	-1,0%	14,1%	0,0%	2,7%	2,0%	0,2%	1,0%	2,3%	0,0%	-0,9%	-1,0%	-3,7%	-2,0%	-9,8%	-1,0%	-9,0%	0,0%	-1,3%	1,0%
SE	0,0%	1,0%	0,0%	2,0%	0,0%	-2,0%	0,0%	3,0%	0,0%	-6,0%	0,0%	3,0%	0,0%	-7,0%	0,0%	3,0%	0,0%	0,0%	0,0%	5,0%	0,0%	-3,0%
UK	8,7%	-2,0%	8,8%	1,0%	4,5%	2,0%	-0,6%	3,0%	2,0%	2,0%	0,3%	1,0%	3,4%	-1,0%	2,2%	-2,0%	2,0%	-1,0%	0,0%	0,0%	-2,9%	1,0%