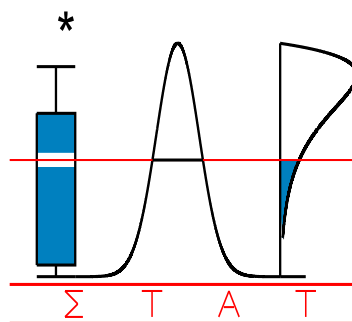# T E C H N I C A L
R E P O R T

0434

TESTING FOR CATCHING-UP: STATISTICAL ANALYSIS OF DEA EFFICIENCY ESTIMATES

HENDERSON, D. J. and V. ZELENYUK

I A P   S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

Testing for Catching-up: Statistical Analysis of DEA Efficiency Estimates[*]

Daniel J. Henderson[†]
Department of Economics
State University of New York at Binghamton

Valentin Zelenyuk[‡]
Institut de Statistique
Université catholique de Louvain, Louvain-la-Neuve

November 28, 2004

Abstract

We use advances in DEA techniques to examine efficiency scores and investigate the issue of convergence/divergence in a cross-country analysis. Specifically, we take use of bootstrapping techniques to examine a data set of 52 developed and developing countries. We find that when using the standard DEA model the results are sometimes less than desirable. Further, we break the sample into the two groups to examine the two-club convergence phenomenon. We find that efficiency scores are significantly different between the two groups and that there is some evidence of convergence of efficiency scores within each group.

Keywords: Efficiency, DEA, Bootstrapping, Convergence
JEL Classification: C14, C15, C44, O1

[†]Corresponding author: Daniel J. Henderson, Department of Economics, State University of New York, Binghamton, NY 13902, (607) 777-4480, Fax: (607) 777-2681, e-mail: djhender@binghamton.edu.

[‡]Valentin Zelenyuk, Institut de Statistique, Université catholique de Louvain, 20 voie du Roman Pays, B 1348, Louvain-la-Neuve, Belgium and UPEG (Ukrainian Productivity and Efficiency Group) at EROC of EERC-Kiev at the National University "Kyiv-Mohyla Academy," zelenyuk@stat.ucl.ac.be.

1 Introduction

Economic growth research has received substantial recognition in recent years. The two major strands being the cross-sectional type regressions found in Baumol (1986), which seek to determine whether there is a tendency for the world's economies to converge over time (poor catching up with the rich), and the decomposition of growth into components attributable to capital deepening and technological progress going back to Solow (1957). However, there is a third strand of research which has become increasingly popular; a method based on Malmquist Productivity Indexes (Caves, Christensen and Diewert 1982), computed via the Data Envelopment Analysis (DEA) estimator. Beginning with Färe, Grosskopf, Norris and Zhang (1994), this strand has introduced a third component into economic growth: efficiency, the ability of a given country to fully exploit available resources it has in producing total output. While much of the mainstream research suggests making adjustments to the input mix (e.g. increasing the level of physical or human capital), if the DEA approach shows that efficiencies are found to affect the growth of labor productivity, then perhaps policymakers should also address methods that would improve efficiency (e.g. establishing macroeconomic and political stability).

Conceptually, the efficiency component is nothing but the residual, somewhat like the 'Solow residual,' that proxies for the aggregated effect of various factors, other than technology and standard inputs on producing total output. This efficiency component can also be understood through Leibenstein's (1966) 'X-efficiency' concept, related to the internal and external motivation of an agent. In our case, X-efficiency would be related to the aggregate result of influence by local and international institutions onto each particular country. X-efficiency is an abstract concept, which of course is unobserved, and in practice is often proxied via the Debreu (1951)-Farrell (1957) measure of technical efficiency, which is usually estimated using the DEA estimator (e.g., see Leibenstein and Maital 1992).

The DEA method of estimating efficiency, however, has received some opposition. There are those who believe that the entire world does not exist under one production frontier. There are still others who believe that DEA has inherent flaws. Traditional (or old paradigm) DEA assumes that no measurement error exists and that the estimated production frontier is piecewise linear. Further, Koop, Osiewalski and Steel

(1999) state that "the sensitivity of DEA to outliers is no doubt one of the weaknesses of the DEA approach. In particular, it is difficult to present some measure of uncertainty (e.g. confidence intervals) using DEA methods." To combat comments such as these, Simar and Wilson (1998, 2000) and others have introduced bootstrapping into the DEA framework. Their methods, based on statistically well-defined models, allow for consistent estimation of the production frontier, corresponding efficiency scores as well as standard errors and confidence intervals. Although advances were made to DEA, these have not been included in many recent papers which examine macroeconomic growth. Recently, Kumar and Russell (2002) employed standard DEA production-frontier methods to analyze convergence by decomposing labor productivity growth into components attributable to technological change, technological catch-up (changes in efficiency) and physical capital accumulation. In their results, they find the main factor driving economic growth to be capital accumulation. Henderson and Russell (2004) extend Kumar and Russell (2002) by adding human capital accumulation into the decomposition and show that about one-third of the productivity growth attributed by Kumar and Russell (2002) to physical capital accumulation should instead be attributed to the accumulation of human capital. Further, they show that the qualitative shift from a unimodal to a bimodal distribution in labor productivity (over a 25 year period) is accounted for primarily by efficiency changes (with physical capital accumulation as a secondary factor – the opposite result of Kumar and Russell 2002). However, both of these papers are subject to the same scrutiny as Färe, Grosskopf, Norris and Zhang (1994). If research is going to continue in this area, it needs to make notice of the advancements in DEA which address the current concerns.[1] In this paper we will focus on circumventing the drawbacks of DEA in an empirical context. Specifically, we will use the recently developed techniques in the statistical analysis of DEA estimates to check for robustness of efficiency estimates for a sample of 52 developed and developing countries. We will also investigate the issue of convergence/divergence in terms of efficiency across countries.

Various empirical studies on economic growth have brought convincing evidence

---

[1]It must be noted that there are other useful approaches to estimating efficiency, with one of the most popular being the stochastic frontier analysis paradigm originated by Aigner, Lovell and Schmidt (1977). For other examples of this approach see Horrace and Schmidt (1996, 2000) and Schmidt and Sickels (1984).

that the world consist of at least these two groups: developed and developing countries. These groups are indeed distinct in their performance as well as in the key factors determining it (especially in institutional development). Quah (1996) has theoretically justified the possibility of the existence of two clubs in the world, with convergence within them and divergence between them, claiming empirical tendency for such phenomenon to be true. In our work, we will employ the notion of 'Catching-up' first discussed in the seminal paper of Abramovitz (1986). Initially envisioning this phenomenon, Abramovitz's argument is based on the discovery of the considerable reduction in the coefficient of variation of growth rates within a group of 16 industrialized countries. Later, Färe, Grosskopf, Norris and Zhang (1994) re-formalized the notion of Catching-up as the decrease overtime in the distance between the actual performance of a country and its potential, according to the best-practice frontier (i.e., as the decrease in inefficiency of the countries over time). In the spirit of Färe, Grosskopf, Norris and Zhang (1994), we will consider three types of Catching-up: (i) within the entire sample, (ii) within distinct groups in the sample, and (iii) between these groups. We have two distinct groups in mind: developed and developing countries.[2]

Specifically, we first use the study of Henderson and Russell (2004) as a stepping-stone and compare our bootstrap bias-corrected efficiency scores with the results of their study. We then break the sample into the two groups to see if the efficiency scores are consistent with the two club convergence hypothesis. The remainder of this paper is organized as follows: Section 2 describes the theory of efficiency measurement and gives a brief description of the current advances in the literature. The third section describes the data, while the fourth section presents the results of the experiment. Finally, Section 5 concludes.

2   Methodology

In this section we discuss the backgrounds of efficiency measurement as well as the latest research advances which will help us obtain more accurate measures for our

---

[2]To check for robustness we have also tried several other ways to group the countries. For example, we analyzed the groups OECD vs. non-OECD as well as defining the developing countries as those from node 4 in Durlauf and Johnson (1995, p. 374) and find that our conclusions do not change; the results can be obtained from the authors upon request.

problem. Although these procedures can be used to analyze any number of (macro or micro) decision making units with multiple inputs and outputs, here we will describe the special case related to our example. For each country i $(i = 1, 2, ..., n)$ we will use the period-t input vector $x_i^t = (K_i^t, H_i^t \times L_i^t)$, where K_i^t is physical capital, and $H_i^t \times L_i^t$ is human capital augmented labor. Further, y_i^t is a single output (GDP) for country i in period t (all inputs and outputs are assumed to be positive). The technology of converting inputs into GDP for each country i, in each time period t, can be characterized by technology set

$$T_i^t \equiv \left\{ \left(x_i^t, y_i^t \right) \mid x_i^t \text{ can produce } y_i^t \right\}.$$

Equivalently, the same technology can be characterized by the output sets

$$P_i^t(x_i^t) \equiv \left\{ y_i^t \mid x_i^t \text{ can produce } y_i^t \right\}, \qquad x_i^t \in \Re_+^2.$$

Here we assume that the technology follows standard regularity assumptions, under which the Shephard (1970) output oriented distance function

$$D_i^t(x_i^t, y_i^t \mid P_i^t(x_i^t)) = \inf \left\{ \theta \mid y_i^t / \theta \in P_i^t(x_i^t) \right\}$$

gives a complete characterization of the technology i in period t, in the sense that we always have

$$D_i^t(x_i^t, y_i^t \mid P_i^t(x_i^t)) \leq 1 \Leftrightarrow y_i^t \in P_i^t(x_i^t).$$

This function (for a single output) is simply the ratio of actual output to the maximal (or potential) output (GDP) that can be produced from the same amount of inputs. A related concept is the Farrell output oriented technical efficiency, defined as

$$TE_i^t \equiv TE_i^t(x_i^t, y_i^t \mid P_i^t(x_i^t)) = \sup \left\{ \theta \mid \theta y_i^t \in P_i^t(x_i^t) \right\} = 1/D_i^t(x_i^t, y_i^t \mid P_i^t(x_i^t)).$$

When $TE_i^t = 1$, the country is considered technically efficient (on the upper boundary of $P_i^t(x_i^t)$ defined as $\partial P_i^t(x_i^t)$), whereas when $TE_i^t \geq 1$ (or ≤ 1 when the commonly used reciprocal is employed), the country is deemed technically inefficient.

Since the true technology and hence output sets are unknown, the individual values of technical efficiency must be estimated. One such popular tool is DEA. This procedure involves measuring given observations $XY^t = \{(x_i^t, y_i^t)\}_{i=1}^n$ against *one* "best practice" frontier (thus sub-script i can be dropped), defined as

$$\partial \widehat{P^t}(x) = \left\{ y \mid y \in \widehat{P^t}(x), \lambda y \notin \widehat{P^t}(x), \lambda \in (1, \infty) \right\},$$

where $\widehat{P}(x)$ is the DEA-estimate of the output set $P(x)$, defined as

$$\widehat{P^t}(x) \equiv \left\{ y \mid \sum_{i=1}^{n} z_i y_i^t \leq y, \sum_{i=1}^{n} z_i x_i^t \leq x, z_i \geq 0, i = 1, 2, ..., n, \sum_{i=1}^{n} z_i = 1 \right\},$$

and where z_i are the intensity variables. The last constraint in the expression above imposes the variable returns to scale (VRS) assumption and its removal would instead impose constant returns to scale (CRS). Geometrically, for the VRS case, the resulting estimate of the best practice technology represents the smallest convex free-disposable hull of the data XY^t. In the theory and measurement of economic growth, the assumption of CRS is usually employed (as well as in Henderson and Russell 2004), and we use it for our study as well, in which case the best practice technology represents the smallest convex free-disposable cone of the data.

Further, the DEA estimate of individual technical efficiency at any fixed point (x, y), is computed relative to the estimated frontier, solving the following linear programming problem:

$$\widehat{TE^t}(x, y | \widehat{P^t}(x)) = \max_{\theta, z_1, z_2, ..., z_n} \left\{ \theta \mid \theta y \in \widehat{P^t}(x) \right\}.$$

Although this procedure is relatively simple, there is a cost of using this DEA estimator. It should be obvious that, when no measurement error exists, that $\widehat{P^t}(x) \subseteq P^t(x)$. Therefore, $\partial \widehat{P^t}(x)$ is a downward biased estimator of $\partial P^t(x)$. As a result, $\widehat{TE^t}(x, y | \widehat{P^t}(x))$ is a downward biased estimate of $TE(x, y | P^t(x))$. In our case, DEA would rate countries as more efficient than they truly are. Even though this bias vanishes asymptotically, in finite samples the bias may be large (especially when the dimension of the DEA model is large relative to the sample size).

Statistical bootstrap procedures have become a useful and popular tool to estimate bias corrected estimates. They also allow for constructing confidence intervals for estimators with unknown sampling distributions, which was impossible for DEA before and was one of the major downsides of this technique. Now we will overcome this problem by using the bootstrap procedures for DEA developed by Simar and Wilson (1998, 2000), Kneip, Simar and Wilson (2003) and Simar and Zelenyuk (2003).

To briefly outline the idea of the bootstrap for individual DEA estimates, let $\mathcal{P}^t$ denote the Data Generating Process (DGP) in period t. It is characterized in terms

5

of the technology set and the distribution of inputs, output and inefficiency across countries in each period (see Kneip, Simar and Wilson 2003). Note that neither the technology, nor any of the distributions are known, and all that is observed is the random sample XY^t generated by $\mathcal{P}^t$. Further note that we have a consistent (DEA) estimate of technical efficiency at any fixed point (x, y). This estimate has some sampling distribution which can be estimated via a consistent bootstrap analog. The DEA estimator of output sets that uses (the randomly sampled with replacement) data XY^{t*} generated from $\widehat{\mathcal{P}^t}$ would then give $\widehat{P^{t*}}(x)$. Further, from the perspective of XY^{t*} this is the true output set, and in our original setting was an estimate of $P^t(x)$. From here, we can calculate a consistent bootstrap estimate of technical efficiency at the same fixed point (x, y) using

$$\widehat{TE^{t*}}(x, y | \widehat{P^{t*}}(x)) = \max_{\theta, z_1, z_2, \dots, z_n} \left\{ \theta \mid \theta y \in \widehat{P^{t*}}(x) \right\}.$$

Intuitively, the unknown distribution of the difference between the true and estimated efficiency score is approximated by the distribution of the difference between the estimated and the bootstrapped efficiency score which in principal is known. This relationship allows consistent estimation of the bias and confidence intervals for $\widehat{TE^t}(x, y | \widehat{P^t}(x))$, at any fixed point (x, y).

In practice, the bootstrap distribution is approximated by generating B samples of XY_b^{t*} $(b = 1, 2, \dots, B)$ drawn from $\widehat{\mathcal{P}^t}$. Then, the technical efficiency is re-estimated B times, yielding $\widehat{TE_b^{t*}}(x, y | \widehat{P_b^{t*}}(x))$. The bootstrap bias estimate and confidence interval for the DEA-estimate of efficiency score $\widehat{TE^t}(x, y | \widehat{P^t}(x))$ can then be obtained from $\widehat{TE_b^{t*}}(x, y | \widehat{P_b^{t*}}(x))$.

Sometimes we do not simply want to look at an individual country versus the others; we may want to examine groups of countries. A naive way of doing this is to compare the sample means of efficiency scores of the two groups. This ignores the relative economic weight of each country in the sample, since all efficiency scores are standardized to be between 1 and infinity. Recently, Färe and Zelenyuk (2003) proposed aggregation of individual efficiencies into *group* efficiencies, where the weights (and the aggregation function) are derived from economic optimization criterion.

To briefly outline this aggregation method, suppose we want to focus on a group within a population, call it group l $(l = 1, \dots, L)$, represented by a sub-sample $i_l = 1, 2, \dots, n_l$ $(\leq n)$ of the original sample of n countries. We will denote the input

6

allocation among firms within group l in period t by $X^{l,t} = (X_1^{l,t}, ..., X_{n_l}^{l,t})$ and the sum of output over the countries in group l in period t with $\overline{Y}^{l,t} = \sum_{i_l}^{n_l} y_{i_l}^{l,t}$. The group technology in period t is defined by assuming a linear structure of aggregation of the output sets. Given this structure of aggregation, it can be shown for the single output case, that the technical efficiency of group l disaggregates exactly into the weighted average of individual technical efficiencies within group l. The weights here being the observed output (i.e. GDP) shares of each country in group l, i.e.,

$$\overline{TE^{l,t}} = \sum_{i_l=1}^{n_l} TE_{i_l}^{l,t}(x_{i_l}^{l,t}, y_{i_l}^{l,t}) \times S_{i_l}^{l,t}, \qquad S_{i_l}^{l,t} = \frac{y_{i_l}^{l,t}}{\overline{Y}^{l,t}} \qquad l = 1, ..., L.$$

Similarly, if we assume the linear structure of aggregation for the output sets over all the (non-overlapping) groups in the population, then the technical efficiency for entire population would disaggregate exactly into the weighted sum of the technical efficiencies for each group, with the weights being the output shares of each group. In other words, this expression tells us that the technical efficiency of the developed and developing countries considered together under the same best-practice frontier can be obtained by averaging the group efficiencies over all groups, with the weights being the shares of GDP for each group in the entire sample (all with respect to the same period).

The aggregate efficiencies outlined above are not observed but can be estimated by replacing the true individual efficiency scores with their consistent DEA estimates. Such estimates of $\overline{TE^{l,t}}$ ($l = 1, ..., L$) and $\overline{TE^t}$ would also fall prey to the same issues as the individual efficiency estimates and a bootstrap procedure can be used to construct the corresponding confidence intervals and to correct for bias (in a similar fashion as outlined above).

Once we have obtained aggregate efficiency scores for two or more groups, we may want to test that these aggregate scores are different from one another. Previously, when making judgement on efficiency of certain groups after using DEA, researchers often resorted to using the Kruskal-Wallis test. The direct application of this test to DEA estimates does not take into account the fact that the estimates are used instead of the true efficiencies. Even more problematic is that such tests use equal weights and ignore the economic weights associated with these standardized efficiency scores. In order to sidestep this problem, we use the test provided in Simar and Zelenyuk

(2003). When one wants to test that the (aggregate) efficiency of two groups (e.g., A and Z) of countries are different from one another, one can test the null that $\overline{TE^{A,t}} = \overline{TE^{Z,t}}$. Specifically, we are interested in whether the relative difference $(RD_{A,Z} \equiv \overline{TE^{A,t}}/\overline{TE^{Z,t}})$ statistic is different from unity. The quantity $RD_{A,Z}$ is not observed but can be estimated by replacing $\overline{TE^{A,t}}$ and $\overline{TE^{Z,t}}$ with their DEA estimates. Further, the bootstrap confidence interval for $RD_{A,Z}$ can be constructed in a similar fashion as discussed above (see Simar and Zelenyuk 2003 for details). The null hypothesis must then be rejected if this confidence interval does not contain unity, and fail to reject otherwise.

3   Data

For comparison purposes we use the same data as Henderson and Russell (2004). For aggregate output we use real GDP per worker (RGDPCH multiplied by POP), whereas physical capital and employment are obtained through the capital per worker variable (KAPW and RGDPW) of which all are obtained from the Penn World Tables Mark 5.6 (focusing on the years 1965 and 1990, and the changes over that 25-year period). For human capital we also use the Hall and Jones (1999) construction, based on the Barro and Lee (2001) average years of education data and the Psacharopoulos (1994) survey of wage equations evaluating the returns to education. Specifically, we determine the level of human capital per country per time period, by letting e_i^t represent the average years of schooling in country i at time t, as

$$H_i^t = \exp(\phi(e_i^t)),$$

where ϕ is a piece-wise linear function, with a zero intercept and a slope of 0.134 through the fourth year of education, 0.101 through the next four years, and 0.068 for any additional years of schooling.

4   Results

4.1   Standard DEA Efficiency Measurement

Tables 1 and 2 list the efficiency scores for each of the 52 countries for the years 1965 and 1990 respectively. The first column of numbers represents whether or not

a country is defined as developed or developing (a value of 1 defines a developed country). The second column of numbers are the individual efficiency scores (standard estimates) which are identical (if allowing for implosion) to the reciprocal of those found in Henderson and Russell (2004). Among the first things to note are the countries that define the world production frontier in each period. For 1965 the frontier defining countries are Argentina, Mauritius, the Netherlands, Sierra Leone, Spain and the United States (note that Mexico is very close to the frontier as well). However, in 1990, the United States has a drop in efficiency by 8.3 percentage points, Spain has decreases by almost 7 percentage points, the Netherlands by about 10 percentage points, and Argentina by about 54 percentage points. The frontier in 1990 is determined solely by Hong Kong, Italy, Mauritius, and Sierra Leone.

It is worth noting the fact that some developing countries, such as Mauritius and Sierra Leone, have full (technical) efficiency and thus define portions of the frontier in both periods. At the same time, most of the developed countries were inefficient with respect to the best practice frontier. This somewhat unexpected result should be interpreted as a failure of DEA. On the opposite, the advantage of the DEA approach is that it allows comparing all types of countries, small and large, poor and rich, with respect to one frontier without imposing any specific parametric functional form.

This artifact that some poor countries turned out to be more efficient than many rich countries should not be misinterpreted: higher efficiency does not imply higher well-being. It only means that countries with higher efficiency scores have exploited their resources relatively better than other countries in the sample with similar levels of inputs (e.g., Malawi, Zimbabwe and Zambia, etc). If those poor but more efficient countries were also able to increase all inputs by the same proportion, while maintaining current efficiency and constant returns to scale, then they would also have higher GDP per worker than the less efficient developed countries.

Thus, what the DEA results suggest to us is that Mauritius and Sierra Leone have exploited all the efficiency (technical and scale) they could, given their resources, and their further economic growth in GDP can come from either changes in technology or in inputs (e.g., physical or human capital accumulation). At this point, however, we would abstain from such a conclusion, until we see the results of the bootstrap-based bias correction DEA estimates.

Finally, Tables 1 and 2 also indicate remarkable increases in efficiency of countries like Hong Kong, South Korea and Thailand. Further, some Western European countries also have improved their efficiency considerably (Belgium, Greece, Ireland and Italy) while some others have become more inefficient relative to the 1990 best-practice frontier (Austria, Denmark and Norway).

4.2   Bootstrapping Individual DEA Efficiency Estimates

The results discussed above are calculated using standard DEA efficiency scores. As stated previously, these estimates posess small sample bias and we use a bootstrap to correct for these biases. In order to perform the Simar and Wilson (1998) smooth homogeneous bootstrap estimation, we have to make an assumption on the density of efficiency scores being independent of the distributions of inputs and output. In addition, we need to incorporate our a priori assumption of the existence of two distinct groups in the population. These groups (while still sharing the same frontier) might have two different distributions of inefficiency. The group-wise heterogeneous DGP, that assumes homogeneity within the groups but allows for heterogeneity between them (see Simar and Zelenyuk 2003), might be more appropriate here. If the homogeneity assumption is correct, then the smooth homogeneous bootstrap would be more efficient than the group-wise heterogeneous analog, but inconsistent if the group-wise heterogeneity assumption is correct. This is a common econometrics efficiency-consistency trade-off, and a statistical test is desirable in order to make this choice. Here, we employ the Simar and Zelenyuk (2004) adapted Li (1996, 1999) test for the equality of densities of two random variables (estimates of efficiency scores). The results of this test are given in Table 3. The test results confirm our earlier suspicions that the distributions of efficiency scores across the groups differ with high probability (for both periods). This justifies the choice of the group-wise heterogenous smooth bootstrap (instead of the homogenous one), which we use to obtain the results presented in the next two tables.

As hypothesized, each of the countries initially considered efficient now fall below the upper boundary of the output set. The average efficiency score in both periods fell from approximately 1.6 to 1.9. Most notable of these frontier defining countries is the case of Sierra Leone. The bootstrap procedure gives a bias corrected efficiency score of

10

1.43 in 1965 with similar results in 1990. This result may solidify the procedure used since many viewed Sierra Leone being on the frontier as a serious case of measurement error (the Penn World Table grades the accuracy of data of each country on an A to F scale, to which Sierra Leone received a D). The bias is corrected for all other countries on the DEA-estimated frontier, including the United States, so that no country has an efficiency score of unity (the ranks, however, are mostly preserved).

4.3 Analysis of Efficiency Distributions

It has become common to use nonparametric kernel density estimation techniques to graphically illustrate various results. This method is also useful in our context since we did not impose distributional assumptions on the efficiency scores across countries. There are, however, several complications which occur when one attempts to do this (see Simar and Zelenyuk 2004). The bottom line is that one has to take special care of at least three things. First, the random variable for which we want to estimate the density has a bounded support, with many observations close to the bound. This bound must be taken into account; we do so by using the Silverman reflection method.

Second, we do not observe the true realizations of the random variable (efficiency) whose density we want to estimate, but only consistent estimates that are biased downward (which would most likely be reflected in biased estimation of the density). Third, some countries (at least one) are on the bound (having an efficiency score equal unity) so there is always a strictly positive probability of observing at least one country with the efficiency score of unity. This is a violation of the continuity assumption that we need to ensure consistency of the density estimation. To take care of this, we follow the suggestion of Simar and Zelenyuk (2004). We first correct for the bias with the bootstrap (as we did in the previous section) and then estimate the densities from the bias corrected efficiency scores. We estimate the densities for each group using the reflection method, with a Gaussian kernel and bandwidth selected via the Sheather and Jones (1991) method. The results are presented in the figure. It provides convincing evidence that the two distributions (non-weighted) are very likely to be different, as we have concluded from the test before. It also support the test results; telling us that the distributions of efficiency scores of countries within each

group (especially for developing countries) seems to have changed minimally over the 25 year period.

The figure (supported by Simar-Zelenyuk adapted Li-test) provides some empirical support for absence of 'Catching-up *between* the groups' (the club of the rich and club of the poor). It is useful to note that while some countries have improved their efficiency (got closer to the best-practice frontier), others have worsened (although overall the distributions remained the same). As a result, there is little support for the 'Catching-up within the groups' hypothesis (for the non-weighted results) from the figure. However, this may be due to the relatively low power of the Li (1996) test, especially for small samples (such as ours). In the next section, we will look for evidence when the economic importance of the standardized efficiency scores are incorporated.

4.4   Aggregate Efficiency

More interesting than the efficiency scores themselves are the conclusions that Henderson and Russell (2004) make about the impact of changes in efficiency on bimodalism. They state that, when accounting for differences in human capital across countries, efficiency changes are the primary driving force in international polarization. The idea behind this two-club or twin-peak convergence is that over the last few decades the distribution of labor productivity has been transformed from a unimodal into a bimodal distribution with a higher mean. In other words, some view the world as becoming divided into the rich and the poor.

To incorporate our knowledge that different groups may have different distributions, we again choose the group-wise heterogeneous version of the sub-sampling bootstrap. The results of the bootstrap on aggregate efficiencies for 1965 and 1990 are displayed in Tables 4 and 5. The eight rows in descending order are: aggregate (weighted[3]) efficiency for the developed countries, aggregate efficiency for the developing countries, the aggregate efficiency scores for all 52 countries, the mean (non-weighted) efficiency for the developed countries, the mean efficiency for the developing countries, and the $\widehat{RD}$ statistic for both the weighted and non-weighted measures.

---

[3]We have also estimated the results by using alternative weighting schemes. For example we used relative population and relative labor. We find that the conclusions of the paper are not significantly different, but the results are available from the authors upon request.

The columns represent: the DEA-estimates, the bias corrected DEA-estimates, the estimated standard errors and the 95%-confidence lower and upper bounds.

A common perception is that developed countries are more efficient than the developing countries. However, as we noted before, being 'rich' does not necessarily imply being efficient in the sense of full exploitation of endowed resources. In fact, it is often the case that being rich still allows some room for efficiency improvement. Nevertheless, the results of Table 4 (1965) do support that common perception that (at least on average) the developed countries are more efficient than the developing countries. This is confirmed by both the weighted and non-weighted averages. Further, the $\widehat{RD}$ statistic is significantly different from unity (i.e., confidence intervals do not overlap with unity). The estimated group efficiency of the developing countries is statistically different from unity for both the weighted and non-weighted aggregations, as suggested by the confidence intervals. This is not the case with the developed countries. According to the estimated confidence intervals for the weighted averages, the efficiency of the developed countries is not statistically different from unity (perfect efficiency). However, it is statistically different from unity when the non-weighted average is used. Thus, if the aggregation of efficiencies that accounts for the relative weight of each country (derived from economic optimization) is preferred, then we cannot reject the hypothesis (at the 5% level) that the group of developed countries is efficient, in the Farrell sense.

Interestingly, the estimated average inefficiency of the entire world is much higher when the non-weighted aggregation is used. This is due to the fact that the weight of the developed countries (the more efficient group) in the sample is 81.5%.

The results for 1990 are placed in Table 5 and although they are quite similar, they have slight but important differences from those in Table 4. Remarkably, the estimated weighted average inefficiency of the developed countries is now larger than it was in 1965. We now reject the null, concluding that the efficiency score for the group of developed countries is significantly different from unity. Again we find the confidence intervals of the $\widehat{RD}$ statistic not overlapping unity, indicating significant difference in efficiencies between the two groups.

It is interesting to examine the dynamics of efficiency for each group of countries. In particular, we would like to find some evidence for presence of the following

phenomena:

i. Efficiency Catching-up/Lagging-behind *within* a group—which we define as a situation when the average (weighted or non-weighted) efficiency for the group is increasing/decreasing.

ii. Efficiency Convergence/Divergence *within* a group—when a measure of 'spread' (such as the coefficient of variation, etc.) for the weighted or non-weighted aggregates is decreasing/increasing.

iii. Efficiency Catching-up/Lagging-behind *between* groups—when the average (weighted or non-weighted) efficiency of any group is getting 'closer/further away' to that of a more efficient group.

iv. Efficiency Convergence/Divergence *between* groups—when a measure of spread for (weighted or non-weighted) efficiencies between groups is decreasing/increasing.

v. *Overall* Efficiency Catching-up—when the average (weighted or non-weighted) efficiency of the entire population is increasing.

vi. *Overall* Efficiency Convergence/Divergence—when a measure of spread for (weighted or non-weighted) efficiencies for the entire population is decreasing/increasing.

At this point, formal statistical tests (based on DEA) for presence of most of the above listed dynamic phenomena do not exist and we are only ready to list a few interesting findings suggested by our results (in the hope to be provocative for further research).

Comparing the results of Tables 4 and 5, we note that according to the bias corrected non-weighted means, the average efficiency of developed countries has not changed much between the periods (1.41 vs. 1.40). However, the inefficiency of the developing countries has slightly increased (2.19 vs. 2.27). This is in line with the results of the density comparison across groups. However, when the economic weight of each country is accounted for (using the weighted aggregation), the conclusions change. Although the weighted average inefficiency of the developing countries has not changed much over the 25 year period, the efficiency of the developed countries

has decreased—giving us some evidence of the efficiency lagging-behind within this group. Further note that there is a slight increase in the value of the $\widehat{RD}$ statistic. This gives slight evidence of Catching-up between the groups, while the decrease in the range of estimated confidence intervals of the $\widehat{RD}$ statistic over time gives some evidence for efficiency convergence between the groups.

The weighted average of the two groups taken together also suggests a decrease in efficiency in the entire sample (mainly because the group of developed countries has the majority economic weight when the two groups are taken together). In other words, we find evidence for further overall efficiency lagging behind across all countries (on average). This is further supported by the fact that (based on the bootstrap confidence intervals) we failed to reject the null that the group of developed countries are perfectly efficient in 1965 but firmly rejected that null in 1990.

5   Conclusion

In this paper we have introduced recent advances in bootstrapping and DEA in order to analyze cross-country efficiency scores and to study the international convergence/divergence hypothesis. We describe the theory behind output based technical efficiency, methods to estimate the efficiencies of groups of countries by their economic weights, as well as methods used to correct for the small sample biases of individual and aggregate DEA estimates. We apply these methods to the study of Henderson and Russell (2004). We find that although many of their results are robust, the efficiency scores of particular countries are sometimes highly misjudged (due to bias) when employing standard DEA techniques. Further, we separate the countries into those that are developed versus those which are developing. As expected, we find that there are significant differences between the two groups in each of the periods (measured with both weighted and non-weighted aggregations). Remarkably, the (non-weighted) distributions of efficiency within these two groups have not changed significantly, according to tests of equality for densities.

As to change across time (over 25 years), our aggregate efficiency indexes suggest some evidence for 'Efficiency Lagging-behind *within*' the group of developed countries (with seemingly no change for the group of developing countries). Interestingly, we also find some evidence for slight 'Efficiency Catching-up *between* the groups' as

well as 'Efficiency Convergence *between* the groups.' At the same time, our indexes suggest the slight presence of 'Efficiency Convergence *within* each group.' Finally, we suggest that a natural extension of our work would be further development of formal statistical tests for the presence of such dynamic phenomena.

References

[1] Abramovitz, M. (1986). "Catching up, forging ahead and falling behind," *Journal of Economic History*, 46, 386-406.

[2] Aigner, D. J., C. A. K. Lovell and P. Schmidt (1977). "Formulation and Estimation of Stochastic Frontier Production Function Models," *Journal of Econometrics*, 6, 21-37.

[3] Barro, R. J. and J-W. Lee (2001). "International Data on Educational Attainment: Updates and Implications," *Oxford Economic Papers*, 53, 541-63.

[4] Baumol, W. J. (1986). "Productivity Growth, Convergence, and Welfare," *American Economic Review*, 76, 1072-85.

[5] Caves, D., L. Christensen and W. E. Diewert (1982). "The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity," *Econometrica* 50, 1393-414.

[6] Durlauf, S. N. and P. A. Johnson (1995). "Multiple Regimes and Cross-Country Growth Behaviour," *Journal of Applied Econometrics*, 10, 365-84.

[7] Debreu, G. (1951). "The Coefficient of Resource Utilization," *Econometrica*, 19, 273-92.

[8] Färe, R., S. Grosskopf, M. Norris, and Z. Zhang (1994). "Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries," *American Economic Review*, 84, 66-83.

[9] Färe, R. and V. Zelenyuk (2003). "On Aggregate Farrell Efficiency Scores," *European Journal of Operational Research*, 146, 615-20.

[10] Farrell, M. J. (1957). "The Measurement of Productive Efficiency," *Journal of the Royal Statistical Society*, Series A, 120, 253-81.

[11] Hall, R. E. and C. I. Jones (1999). "Why Do Some Countries Produce So Much More Output Per Worker Than Others?" *Quarterly Journal of Economics*, 114, 83-116.

[12] Henderson, D. J. and R. R. Russell (2004). "Human Capital and Convergence: A Production-Frontier Approach," forthcoming, *International Economic Review.*

[13] Horrace, W. C. and P. Schmidt (1996). "Confidence Statements for Efficiency Estimates from Stochastic Frontier Models," *Journal of Productivity Analysis*, 7, 257-282.

[14] Horrace, W. C. and P. Schmidt (2000). "Multiple Comparisons with the Best, with Economic Applications," *Journal of Applied Econometrics*, 15, 1-26.

[15] Kneip, A., B. Park and L. Simar (1998). "A Note on the Convergence of Nonparametric DEA Estimators for Production Efficiency Scores," *Econometric Theory*, 14, 783-93.

[16] Kneip, A., L. Simar and P. Wilson (2003). "Asymptotics for DEA Estimators in Non-parametric Frontier Models," Discussion Paper, Institut de Statistique, Université Catholique de Louvain, Belgium.

[17] Koop, G., J. Osiewalski, and M. F. J. Steel (1999). "The Components of Output Growth: A Stochastic Frontier Analysis," *Oxford Bulletin of Economics and Statistics*, 61, 455-87.

[18] Kumar, S. and R. R. Russell (2002). "Technological Change, Technological Catch-up, and Capital Deepening: Relative Contributions to Growth and Convergence," *American Economic Review*, 92, 527-48.

[19] Li, Q. (1996) "Nonparametric Testing of Closeness between Two Unknown Distributions," *Econometric Reviews*, 15, 261-74.

[20] Leibenstein, H. (1966). "Allocative Efficiency vs. 'X-Efficiency'," *American Economic Review* 56, 392-415.

[21] Leibenstein, H. and S. Maital (1966). "Empirical Estimation and Partitioning of X-Inefficiency: A Data-Envelopment Approach," *American Economic Review* 82, 428-33.

[22] Psacharopoulos, G. (1994). "Returns to Investment in Education: A Global Update," *World Development*, 22, 1325-43.

18

[23] Quah, D. (1996). "Twin Peaks: Growth Convergence in Models of Distribution Dynamics," *Economic Journal*, 106, 1045-55.

[24] Schmidt, P. and R. C. Sickels (1984). "Production Frontiers and Panel Data," *Journal of Business and Economic Statistics*, 2, 367-74.

[25] Shephard, R. W. (1970). *Theory of Cost and Production Functions*, Princeton, Princeton University Press.

[26] Sheather, S. and M. Jones (1991). "A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation," *Journal of the Royal Statistical Society*, Series B, 53, 683-90.

[27] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

[28] Simar, L. and P. Wilson (1998). "Sensitivity of Efficiency Scores : How to Bootstrap in Nonparametric Frontier Models," *Management Science* 44, 49-61.

[29] Simar, L. and P. Wilson (2000). "A General Methodology for Bootstrapping in Nonparametric Frontier Models," *Journal of Applied Statistics*, 27, 779-802.

[30] Simar, L. and V. Zelenyuk (2003). "Statistical Inference for Aggregates of Farrell-type Efficiencies," Discussion Paper 0324, Institut de Statistique, Université Catholique de Louvain, Belgium.

[31] Simar, L. and V. Zelenyuk (2004). "On Testing Equality of Two Distribution Functions of Efficiency Scores Estimated via DEA," (in progress), Institut de Statistique, Université Catholique de Louvain, Belgium.

[32] Solow, R. W. (1957). "Technical Change and the Aggregate Production Function," *The Review of Economics and Statistics*, 39, 312-20.

Table 1 – Original and bias-corrected DEA estimates (1965)[4]

Country	D	Efficiency	Corrected	Std Error	Lower	Upper
Argentina	0	1.00	1.13	0.03	1.09	1.20
Australia	1	1.35	1.51	0.04	1.45	1.61
Austria	1	1.25	1.41	0.04	1.34	1.49
Belgium	1	1.39	1.56	0.04	1.49	1.65
Bolivia	0	1.99	2.29	0.07	2.18	2.44
Canada	1	1.18	1.33	0.04	1.27	1.41
Chile	0	1.17	1.33	0.03	1.27	1.40
Colombia	0	2.08	2.33	0.05	2.24	2.44
Denmark	1	1.37	1.53	0.04	1.47	1.61
Dominican Rep.	0	1.25	1.46	0.05	1.38	1.58
Ecuador	0	2.37	2.65	0.07	2.54	2.80
Finland	1	1.51	1.69	0.05	1.62	1.80
France	1	1.18	1.32	0.04	1.27	1.40
Germany, W.	1	1.45	1.62	0.04	1.55	1.71
Greece	1	1.81	2.04	0.06	1.94	2.16
Guatemala	0	1.04	1.20	0.03	1.13	1.27
Honduras	0	1.93	2.19	0.06	2.09	2.31
Hong Kong	0	2.17	2.45	0.07	2.33	2.59
Iceland	1	1.06	1.19	0.03	1.14	1.26
India	0	2.28	2.78	0.14	2.56	3.09
Ireland	1	1.49	1.68	0.05	1.60	1.77
Israel	1	1.63	1.82	0.04	1.75	1.92
Italy	1	1.32	1.48	0.04	1.41	1.57
Jamaica	0	1.61	1.82	0.05	1.74	1.93
Japan	1	1.83	2.08	0.06	1.98	2.20
Kenya	0	3.20	3.82	0.15	3.58	4.18
Korea, Rep.	0	2.41	2.84	0.10	2.68	3.06
Malawi	0	3.69	5.03	0.48	4.22	5.94
Mauritius	0	1.00	1.16	0.04	1.11	1.24
Mexico	0	1.00	1.12	0.03	1.08	1.18
Netherlands	1	1.00	1.12	0.03	1.07	1.19
New Zealand	1	1.20	1.35	0.04	1.29	1.43
Norway	1	1.27	1.42	0.04	1.36	1.50
Panama	0	2.17	2.44	0.07	2.33	2.59
Paraguay	0	1.02	1.35	0.12	1.16	1.59
Peru	0	1.52	1.70	0.04	1.64	1.79

[4]We use 2000 group-wise heterogenous bootstrap iterations, a Gaussian kernel, as well as the Silverman (1986) reflection method (to account for boundary problem) and select the bandwidth via the Sheather and Jones (1991) method.

Table 1 (Continued)

Country	D	Efficiency	Corrected	Std Error	Lower	Upper
Philippines	0	2.39	2.76	0.08	2.63	2.94
Portugal	1	1.33	1.51	0.04	1.44	1.59
Sierra Leone	0	1.00	1.43	0.14	1.16	1.68
Spain	1	1.00	1.13	0.03	1.07	1.19
Sri Lanka	0	3.01	3.40	0.09	3.26	3.60
Sweden	1	1.20	1.34	0.04	1.28	1.42
Switzerland	1	1.04	1.16	0.03	1.11	1.23
Syria	0	1.61	1.81	0.05	1.73	1.92
Taiwan	0	1.91	2.18	0.06	2.07	2.31
Thailand	0	2.22	2.71	0.14	2.50	3.02
Turkey	1	1.77	2.02	0.06	1.91	2.14
U.K.	1	1.08	1.22	0.03	1.16	1.29
U.S.A.	1	1.00	1.13	0.03	1.08	1.19
Yugoslavia	0	1.53	1.79	0.06	1.70	1.92
Zambia	0	2.08	2.38	0.07	2.26	2.52
Zimbabwe	0	4.80	5.37	0.12	5.16	5.63
Mean	0.46	1.68	1.93	0.07	1.82	2.07

Table 2 – Original and bias-corrected DEA estimates (1990)[5]

Country	D	Efficiency	Corrected	Std Error	Lower	Upper
Argentina	0	1.54	1.83	0.08	1.67	1.97
Australia	1	1.30	1.43	0.03	1.38	1.49
Austria	1	1.32	1.46	0.03	1.41	1.53
Belgium	1	1.16	1.28	0.03	1.24	1.35
Bolivia	0	2.29	2.75	0.10	2.58	2.95
Canada	1	1.24	1.38	0.03	1.33	1.45
Chile	0	1.54	1.82	0.07	1.68	1.96
Colombia	0	1.76	1.98	0.05	1.90	2.09
Denmark	1	1.46	1.61	0.03	1.56	1.68
Dominican Rep.	0	1.78	2.13	0.08	1.99	2.29
Ecuador	0	2.38	2.66	0.07	2.55	2.81
Finland	1	1.50	1.68	0.04	1.61	1.78
France	1	1.15	1.29	0.03	1.24	1.36
Germany, W.	1	1.35	1.51	0.04	1.45	1.60
Greece	1	1.59	1.75	0.03	1.70	1.83
Guatemala	0	1.14	1.38	0.05	1.30	1.48
Honduras	0	2.18	2.64	0.09	2.49	2.83
Hong Kong	0	1.00	1.19	0.04	1.11	1.28
Iceland	1	1.11	1.23	0.03	1.18	1.29
India	0	2.11	2.66	0.12	2.46	2.91
Ireland	1	1.17	1.30	0.03	1.25	1.37
Israel	1	1.20	1.34	0.03	1.29	1.41
Italy	1	1.00	1.13	0.03	1.08	1.19
Jamaica	0	1.83	2.25	0.08	2.12	2.42
Japan	1	1.66	1.83	0.04	1.77	1.91
Kenya	0	2.68	3.53	0.22	3.14	3.97
Korea, Rep.	0	1.67	1.91	0.05	1.82	2.03
Malawi	0	2.96	4.04	0.30	3.50	4.62
Mauritius	0	1.00	1.23	0.04	1.16	1.33
Mexico	0	1.15	1.33	0.04	1.26	1.40
Netherlands	1	1.10	1.21	0.02	1.18	1.26
New Zealand	1	1.49	1.64	0.03	1.59	1.71
Norway	1	1.54	1.73	0.05	1.65	1.83
Panama	0	2.83	3.26	0.09	3.09	3.45
Paraguay	0	1.00	1.36	0.09	1.18	1.54
Peru	0	2.41	2.86	0.12	2.63	3.07

[5]We use 2000 group-wise heterogenous bootstrap iterations, a Gaussian kernel, as well as the Silverman (1986) reflection method (to account for boundary problem) and select the bandwidth via the Sheather and Jones (1991) method.

Table 2 (Continued)

Country	D	Efficiency	Corrected	Std Error	Lower	Upper
Philippines	0	2.30	2.87	0.11	2.68	3.10
Portugal	1	1.04	1.18	0.03	1.12	1.25
Sierra Leone	0	1.00	1.43	0.13	1.18	1.65
Spain	1	1.07	1.18	0.02	1.14	1.24
Sri Lanka	0	2.84	3.37	0.14	3.09	3.62
Sweden	1	1.39	1.55	0.04	1.49	1.63
Switzerland	1	1.29	1.44	0.04	1.38	1.52
Syria	0	1.20	1.34	0.03	1.29	1.40
Taiwan	0	1.58	1.74	0.03	1.69	1.81
Thailand	0	1.71	2.08	0.07	1.96	2.23
Turkey	1	1.62	1.92	0.08	1.75	2.06
U.K.	1	1.05	1.17	0.03	1.12	1.24
U.S.A.	1	1.08	1.19	0.02	1.16	1.24
Yugoslavia	0	1.73	2.06	0.08	1.92	2.21
Zambia	0	3.01	3.89	0.22	3.52	4.34
Zimbabwe	0	3.94	4.82	0.17	4.54	5.18
Mean	0.46	1.64	1.93	0.07	1.82	2.08

Table 3 - Simar-Zelenyuk-adapted-Li Test for equality of efficiency distributions across groups[6]

Null Hypothesis	Test Statistic	Bootstrap p-value
$f(\text{developed eff}_{65}) = f(\text{developing eff}_{65})$	5.10	0.00
$f(\text{developed eff}_{90}) = f(\text{developing eff}_{90})$	5.06	0.00
$f(\text{developed eff}_{65}) = f(\text{developed eff}_{90})$	-0.20	0.77
$f(\text{developing eff}_{65}) = f(\text{developing eff}_{90})$	-0.30	0.67

---

[6]The number of bootstrap iterations is 2000. For these tests, we use the Gaussian kernel and Silverman (1986) adaptive bandwidth estimator, $h = 0.9An^{-1/5}$ (where $A = \min(\sqrt{var(u)}, iqr(u)/1.349)$, where $iqr(u)$ is the interquartile range of random variable u, which density is estimated.)

Table 4 - Group-Wise Heterogeneous Sub-Sampling Bootstrap for Aggregate Efficiencies (1965)[7]

	DEA Est.	Bias Corrected	Std Error	Lower	Upper
AgEf. Developed	1.19	1.22	0.13	0.90	1.37
AgEf. Developing	1.83	2.08	0.33	1.45	2.55
AgEf. of All	1.30	1.35	0.15	0.99	1.54
MeEf Developed	1.32	1.41	0.08	1.26	1.55
MeEf Developing	1.98	2.19	0.25	1.65	2.60
MeEf of All	1.68	1.84	0.14	1.54	2.07
$\widehat{RD}_{A,Z}$ for AgEff.	0.65	0.53	0.18	0.16	0.82
$\widehat{RD}_{A,Z}$ for MeEff.	0.67	0.62	0.11	0.41	0.82

---

[7]AgEf. = aggregate efficiency (weighted average), MeEf = mean efficiency (non-weighted average). Confidence Intervals are all at the 95% level; sub-sample size in each bootstrap replication is determined via $m_l = n_l^\kappa$, $\kappa = 0.7$.

Table 5 - Group-Wise Heterogeneous Sub-Sampling Bootstrap for Aggregate Efficiencies (1990)[8]

	DEA Est.	Bias Corrected	Std Error	Lower	Upper
AgEf. Developed	1.23	1.31	0.11	1.06	1.44
AgEf. Developing	1.78	2.08	0.22	1.63	2.49
AgEf. of All	1.34	1.47	0.12	1.20	1.64
MeEf. Developed	1.29	1.40	0.07	1.25	1.51
MeEf. Developing	1.95	2.27	0.21	1.85	2.66
MeEf. of All	1.64	1.87	0.12	1.64	2.10
$\widehat{RD}_{A,Z}$ for AgEff.	0.69	0.59	0.14	0.27	0.82
$\widehat{RD}_{A,Z}$ for MeEff.	0.66	0.59	0.10	0.36	0.76

---

[8]AgEf. = aggregate efficiency (weighted average), MeEf = mean efficiency (non-weighted average). Confidence Intervals are all at the 95% level; sub-sample size in each bootstrap replication is determined via $m_l = n_l^\kappa$, $\kappa = 0.7$.

pdf of TE, f(TE)

Farrell output oriented technical efficiency score, TE

Legend:
- Developed in 1990
- Developing in 1990
- Developed in 1965
- Developing in 1965