



HAL
open science

Small is not that beautiful after all: Measuring the scale efficiency of Tunisian High Schools using a DEA-bootstrap method

Hédi Essid, Pierre Ouellette, Stephane Vigeant

► To cite this version:

Hédi Essid, Pierre Ouellette, Stephane Vigeant. Small is not that beautiful after all: Measuring the scale efficiency of Tunisian High Schools using a DEA-bootstrap method. *Applied Economics*, 2011, 45 (09), pp.1109-1120. 10.1080/00036846.2011.613795 . hal-00762811

HAL Id: hal-00762811

<https://hal.science/hal-00762811>

Submitted on 8 Dec 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Small is not that beautiful after all: Measuring the scale efficiency of Tunisian High Schools using a DEA-bootstrap method

Journal:	<i>Applied Economics</i>
Manuscript ID:	APE-2011-0073
Journal Selection:	Applied Economics
Date Submitted by the Author:	17-Feb-2011
Complete List of Authors:	Essid, Hédi; Institut Supérieur de Gestion Ouellette, Pierre; Université du Québec à Montréal, Economics Vigeant, Stephane; Université des sciences et technologie de Lille, Institut des sciences économiques et de management
JEL Code:	D24 - Production Capital and Total Factor Productivity Capacity < D2 - Production and Organizations < D - Microeconomics, I21 - Analysis of Education < I2 - Education < I - Health, Education, and Welfare
Keywords:	Returns to scale, Data envelopment analysis, Bootstrap, Quasi-fixed inputs, Optimal School Size

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

SCHOLARONE™
Manuscripts

For Peer Review

1
2
3
4
5
6
7
8
9 **Small is not that beautiful after all: Measuring the scale efficiency of Tunisian**
10 **High Schools using a DEA-bootstrap method**
11

12 Hédi Essid

13 Institut Supérieur de Gestion
14 41, Rue de la liberté, Cité Bouchoucha 2000
15 Le Bardo, Tunis-Tunisie
16 e-mail: hedi.essid@ihec.rnu.tn
17
18

19
20
21 Pierre Ouellette

22 Department of Economics,
23 Université du Québec à Montréal
24 PO Box 8888, Station Centre-Ville
25 Montréal, Québec
26 Canada H3C 3P8
27 e-mail: ouellette.pierre@uqam.ca
28
29

30
31 and

32 Stéphane Vigeant (corresponding author)
33 IESEG School of management

34 and

35 EQUIPPE (Universités de Lille)
36 Faculté des Sciences Économiques et Sociales
37 Université des Sciences et Technologies de Lille
38 59655 Villeneuve d'Ascq Cédex,
39 France
40

41 Phone: +33 (0)3 20 33 63 56, Fax: +33 (0)3 20 43 66 55
42 e-mail: Stephane.Vigeant@univ-lille1.fr
43
44
45
46
47
48
49
50
51

52 February 2011
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9 **Abstract:** Allocation of resources to schools in a centrally managed State system, as the Tunisian
10 one, should depend on the performance of the individual institutions. The optimal size is of
11 crucial importance in this context and we need accurate measurement for sound policies. This
12 paper discusses and implements a nonparametric statistical test procedure for organization scale
13 efficiency. This procedure allows us to test whether the observed scale efficiency is optimal or
14 not using a smooth bootstrap methodology for efficiency measures estimated using DEA
15 methods. Because school principals do not control for the size of their institution, i.e. the capital
16 available at decision time, the scale efficiency measures are defined so as to include quasi-fixed
17 inputs. The results show that scale efficiency measures are subject to sampling variation. We also
18 found that the schools that are scale efficient are usually mid-sized and large schools, when size
19 is measured by the number of students. This contradicts the largely shared view among decision
20 makers that small schools were optimal.
21
22
23
24
25
26
27
28
29
30
31

32 **Keywords:** Returns to scale, Data envelopment analysis, Bootstrap, Quasi-fixed inputs, Tunisia
33 School Efficiency, Optimal School Size.
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1. Introduction

Public education systems, like the Tunisian one, face scarce resources that have to be shared between educational institutions. One of the objectives of a well designed educational policy is to allocate these resources to the right number of schools. It implies that identifying the optimal size of the institutions is crucial for the organization of the education system. But this debate raises some arguments by proponents of different views of the system. Smaller institutions are said to be able to create a motivating environment and a comfortable sense of proximity for the students, e.g. Coladarci and Cobb (1996) or Monk (1987), while larger institutions have the advantage of gathering together a large amount of resource and offering an optimal mix of services, e.g. Kuziemko (2006). In fact, arguments can be made for both types of school as it is shown in the literature review by Leithwood and Jantzi (2009).

Costs have been central to the analysis of schools optimal size. Most of the earlier studies focused on concepts of average cost. Morris (1964) shows that the cost per student is the lowest in high school with less than 500 students while Turner and Thrasher (1970) find that cost improvements are very slim beyond 1000 students in the same institution. Monk (1987) shows that beyond 400 students, there are no rooms for economies of scale. These results are consistent with the view that the average cost function of schools must be U-shaped (Fox, 1981). McGuire (1989) finds results consistent with an increasing average cost for schools that reach a certain size (beyond 2,000 students, the unit costs starts to be increasing). More recently, Colgrave and Giles (2005) concluded that the optimal efficient scale for a high school is 1,540 students, while Foreman-Peck and Foreman-Peck (2006) are in the range of 540 students. It is also common to find a negative relationship between the average cost and the size for school with less than 1,000 students (e.g. Stiefel *et al.*, 2009, Chabotar, 1989, and Kumar 1983).

In this paper, we offer an element of response to this debate by addressing the question of the qualitative measurement of returns to scale in organizations. This measure allows us to determine if organizations' activities are characterized by either increasing, decreasing or constant returns to scale. Based on this assessment, we can determine the optimal size of a school. The methodology is applied to a sample of Tunisian high school and we use it to give some insight on the way education policies might be oriented.

1
2
3
4
5
6 Our approach is anchored in the DEA tradition. Returns to scale have received substantial
7 attention, leading to three measurement methods under the DEA approach: the scale efficiency
8 method (Färe and Grosskopf (1985) and Färe *et al.* (1994)), the sum of intensity variables method
9 of Banker (1984), Banker and Thrall (1992) and Färe and Grosskopf (1994), and the dual
10 variable-sign-method of Banker *et al.* (1984). Banker *et al.* (1994) have shown that these three
11 methods are equivalent. It is not uncommon however, that practitioners using these methods
12 neglect the statistical content of the efficiency measures and do not distinguish between the true
13 value and the estimator.
14
15
16
17
18
19

20 If the data are generated by a distribution with bounded support, based on the real
21 production set, then the scale efficiency is measured with respect to an estimated frontier with
22 data obtained from an unobserved Data Generating Process (DGP). Under this framework, the
23 observed organization's efficiency is a statistical estimate potentially noisy and spoiled by
24 statistical noises. That is, efficiency measurements are subject to sampling variations.
25 Consequently, it is necessary to adopt a formal statistical methodology that differentiates between
26 the real efficiency value and its estimator in order to evaluate the statistical significance of the
27 results concerning the returns to scale of an organization. This is what the procedure developed
28 by Simar and Wilson (2002) offers. Their method consists in smoothing the probability
29 distribution of the efficiency scores with a convolution kernel. Then the authors use a reflection
30 method (Silverman, 1986 and Schuster, 1985) to correct the estimator of the probability density
31 in the neighborhood of the boundaries. However, Essid *et al.* (2010b) show that this test
32 procedure rests on a statistic that can assume infeasible values. That is, the test is defined as the
33 ratio of two efficiency scores, those being measured under two different assumptions on the
34 return to scale. Then, in the bootstrap simulations, the efficiency scores are generated from two
35 different DGPs. This might ultimately lead to values of the score that are not admissible: The
36 scale efficiency pseudo-scores may end up taking values above one and this is clearly not
37 allowed, by definition.
38
39
40
41
42
43
44
45
46
47
48
49
50

51 Essid *et al.* (2010b) have proposed a solution to this drawback of an otherwise working
52 procedure. They have developed a smooth bootstrap method that ensures that the pseudo-scores
53 of the efficiency scale assume only admissible values and they base the test procedure on a
54 consistent DGP of the scale efficiencies. Furthermore, instead of considering a global test for
55 returns to scale, a test specific to each decision making unit (DMU) is implemented.
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

The methodology used in this paper to measure and test the scale efficiency of Tunisian high school is based on Essid *et al.* (2010b). We have adapted this method to take into account the fact the school principals do not control the physical size of their institution. This means that we include quasi-fixed inputs (also known as non-discretionary inputs) in the DEA model as it was done in Banker and Morey (1986).

The Tunisian high school network offers a perfect study case. Firstly, the education system is state run and resource allocation is centralized. The Department of education creates and implements the programs, it hires the teachers and administrative staff and dispatch them based on the estimated needs of the schools and finally it allocates the operating budget between the different institutions. (The share of the money from the private sector is minimal.) This is to ensure that all schools have more or less the same resources per student. Secondly, the size of the schools measured by the number of students varies a lot over the country. This offers a perfect environment to study the correlation between the size of schools and their returns to scale characterizations. Thirdly, policy makers in the Tunisian Department of Education strongly believe that smaller the institution is, better the performance will be. This raises the problem of the optimal size of the institution and of the scale characterization that schools need to have to profit from the scale efficiency.

This paper offers a test of the scale efficiency of high schools in Tunisia and we check on the claim made by the department of education decision makers that smaller schools are required. We also check the robustness of our results with respect to the simulation methods used. This is done by comparing the results obtained under a homogenous bootstrap procedure to those obtained using a heterogeneous bootstrap procedure.

2. Production technology and returns to scale

Consider a production activity that uses two sets of inputs, one containing those under the direct control of the decision maker, $x = \{x_{i_1}, i_1 = 1, \dots, m_1\}$, and one containing those not under her control at decision time, $z = \{z_{i_2}, i_2 = 1, \dots, m_2\}$, to produce the output vector, $y = \{y_r, r = 1, \dots, s\}$.

The production possibility set of this activity is defined as:

$$\Psi = \{(x, z, y) \in \mathbb{R}_+^{m_1+m_2+s} \mid (x, z, y) \text{ is feasible}\}. \quad (1)$$

We suppose that the production set Ψ is closed, satisfies free disposal of inputs, is bounded for finite inputs and positive outputs require positive inputs. Except for our inclusion of quasi-fixed inputs in the production process, these are standard assumptions on the technology and are discussed in Färe (1988), among others.

To characterize the returns to scale environment of the firms and to test for their type, specific production sets are required. The returns to scale are characterized by the way one can either expand the scale in the production set or shrink it or both. That is, the technology exhibits non increasing (*nirs*), non decreasing (*ndrs*) or constant (*crs*) returns to scale depending on the value assumed by the positive scalar α in the following set:

$$\Psi^k = \{(x, z, y) \in \Psi \mid (\alpha x, \alpha z, \alpha y) \in \Psi \text{ for all } \alpha \in \mathbb{K}^k\}, \text{ for } k = \textit{nirs}, \textit{ndrs}, \textit{crs} \quad (2)$$

where $\mathbb{K}^{\textit{nirs}} = [0, 1)$, $\mathbb{K}^{\textit{ndrs}} = [1, \infty)$ and $\mathbb{K}^{\textit{crs}} = [0, \infty)$. A technology that exhibits *ndrs*, *nirs* or *crs* in different regions of the production frontier is said to be characterized by variable returns to scale (*vrs*). This production possibility set is denoted $\Psi^{\textit{vrs}}$.

It is possible to define an input oriented technical efficiency measure in the sense of Farrell (1957) with respect to the various assumptions concerning the returns to scale. That is:

$$\theta^k(x, z, y) = \min\{\theta \mid (\theta x, z, y) \in \Psi^k\}, \text{ where } k = \textit{nirs}, \textit{ndrs}, \textit{crs} \text{ and } \textit{vrs}. \quad (3)$$

The scalar $\theta(x, z, y)$ is a radial measure of efficiency that gives the maximal factor by which we can reduce x so that y can still be produced, given the quasi-fixed input vector z .

From Färe and Grosskopf (1985), the (technical) efficiency measures defined in (3) can be used to construct scale efficiency measures for each organization. A scale efficiency measure is the ratio of the efficiency measure under *crs* technology and a *vrs* type technology. That is:

$$S_1(x, z, y) = \frac{\theta^{\textit{crs}}(x, z, y)}{\theta^{\textit{vrs}}(x, z, y)} \leq 1. \quad (4)$$

We say that the production technology is of the *crs* type if $S_1(x, z, y) = 1$. To determine the returns to scale of the technology when $S_1(x, z, y) < 1$ we compute a second ratio that is less restrictive than the initial ratio. That is:

$$S_2(x, z, y) = \frac{\theta^{nirs}(x, z, y)}{\theta^{vrs}(x, z, y)} \leq 1. \quad (5)$$

When $S_2(x, z, y) = 1$ the technology is said to exhibit decreasing returns to scale and for values strictly less than one, i.e. $S_2(x, z, y) < 1$, it is said to exhibit increasing returns to scale.

3. Efficiency estimation and statistical model

The description of the technology presented in the previous section characterizes the true unobserved model. This means that the frontier and the scale efficiency measures $S_1(x, z, y)$ and $S_2(x, z, y)$ are not observed. These quantities must be estimated using data.

Consider a sample of n observations, $\Psi_0 = \{(x_j, z_j, y_j)\}_{j=1}^n$. Since the true production set Ψ is not observed, the efficiency score, defined as the gap between the frontier of that set and the units observed performance, is not available. Consequently, the best we can do is to estimate it from the sample, Ψ_0 , by estimating the “missing” frontier of the production set. To do this, we use the convex hull of the sample. The smallest convex envelop of the data gives the DEA estimator in the *vrs* case. That is:

$$\hat{\Psi}^{vrs} = \left\{ (x, z, y) \in \mathbb{R}^{m_1+m_2+s} \mid x \geq \sum_{j=1}^{j=n} \lambda_j x_j, z \geq \sum_{j=1}^{j=n} \lambda_j z_j, y \leq \sum_{j=1}^{j=n} \lambda_j y_j, \sum_{j=1}^{j=n} \lambda_j = 1 \forall j = 1, \dots, n \right\}. \quad (6)$$

To obtain the estimators for the other types of returns to scale, $\hat{\Psi}^{nirs}$, $\hat{\Psi}^{ndrs}$ and $\hat{\Psi}^{crs}$ it is sufficient to slightly alter the constraint on the sum of the λ_j . That is, in the *nirs* case we have $\sum_{j=1}^{j=n} \lambda_j \leq 1$, and in the *ndrs* $\sum_{j=1}^{j=n} \lambda_j \geq 1$ must hold and finally, to obtain a *crs* envelop, no constraints on the λ_j are necessary other than non negativity of the λ s.

Farrell’s efficiency measures are obtained by substituting $\hat{\Psi}^k$ to Ψ^k in equations (3). This gives:

$$\hat{\theta}^k(x, z, y) = \min \left\{ \theta \mid (\theta x, z, y) \in \hat{\Psi}^k \right\} \quad k=vrs, nirs, ndrs, crs. \quad (7)$$

The resulting scale ratio estimators are $\hat{S}_1 = \hat{\theta}^{crs}(x, z, y) / \hat{\theta}^{vrs}(x, z, y)$ and $\hat{S}_2(x, z, y) = \hat{\theta}^{nirs}(x, z, y) / \hat{\theta}^{vrs}(x, z, y)$.

To insure the consistency of the estimator, it is necessary to have an explicit understanding of the way the data are generated. In other words, we have to specify a statistical model that allows us for a full to characterization of the data generating process (DGP). This is the content of the following assumption.

Assumption A1: The set of observations $\{(x_j, z_j, y_j)\}_{j=1}^n$ are identically and independently distributed (*i.i.d.*) random variables with probability density function $f(x, z, y)$ defined on Ψ .

Because Farrell's efficiency measure is radial and because the vector of input x has an equivalent polar representation to the usual Cartesian representation, we have three equivalent representation of the input-output combination. That is:

$$(x, z, y) \Leftrightarrow (\omega, \eta, z, y) \Leftrightarrow (\theta, \eta, z, y), \quad (8)$$

where $\omega(x) = \|x\| = \sqrt{x^T x}$, $\eta = (\eta_1, \dots, \eta_i, \dots, \eta_{m-1})$ and $\eta = \eta(x) \in [0, \pi/2]^{m-1}$ is the angle. Using the law of conditional probability, the joint probability density of (θ, η, z, y) can be decomposed to obtain:

$$f(\theta, \eta, z, y) = f(\theta | \eta, z, y) f(\eta | z, y) f(z | y) f(y). \quad (9)$$

Assumption A2: The probability density function $f(x, z, y)$ is continuous on the interior of Ψ , and $f(x^\partial(z, y), z, y) > 0$ where $x^\partial(z, y) = \theta(x, z, y)x$ is any point on the frontier of Ψ .

Note that points that are not feasible, i.e. the points not in Ψ , have zero probability and the density is identically equal to zero.

Assumption A3: The efficiency measure $\theta(x, z, y)$ is differentiable in x , y , and z .

This assumption gives the necessary smoothness of the frontier function. Kneip *et al.* (1998) have shown that the convergence rate of the estimator depends on this smoothing condition. However, the smoothing condition we have used is slightly stronger than the one

required for convergence. As shown by Kneip *et al.* (1998), it is sufficient that the efficiency measure satisfy a Lipschitz condition.

Assumption A1, A2 and A3 together define the statistical model that allows us to characterize the DGP, denoted \mathfrak{S} . In fact, the DGP is entirely characterized by the production possibility set Ψ and the density f . That is, $\mathfrak{S} = (\Psi, f)$.

This statistical model ensures the convergence of the estimator and permits a characterization of the asymptotic properties of nonparametric estimators of the DEA type. In particular, Kneip *et al.* (1998) have shown the consistency of the DEA estimator in the multidimensional case. They show that for a *vrs* technology with m inputs, the rate of convergence is of order $n^{-2/(m+s+1)}$ where n is the number of DMU. Park *et al.* (2010) show that for a *crs* technology, the convergence rate of the DEA estimator is faster, as it is of the order $n^{-2/(m+s)}$. These results do not make the difference between discretionary and non discretionary inputs. However, Essid *et al.* (2010c) show that, as long as the efficiency measure is radial, the convergence of the estimator is kept independently of the returns to scale the technology exhibits. An important point made is that since it is a nonparametric estimator, the convergence rate is subject to the curse of dimensionality, and the rate of convergence decrease rapidly with the number of quasi-fixed inputs. Bootstrap simulation procedures are used to construct confidence intervals, to correct the estimator's bias and to test statistical hypothesis, as it is now the norm.

4. Hypothesis testing and returns to scale

The traditional approach to qualitative returns to scale measures starts from Färe and Grosskopf (1985). The central argument of the approach is that the ratio of efficiency measures calculated under different returns to scale assumptions, $\hat{S}_{ij} = \hat{\theta}^{crs}(x_j, z_j, y_j) / \hat{\theta}^{vrs}(x_j, z_j, y_j) \leq 1$, reveal some information on the nature of the returns to scale of the DMU. When $\hat{S}_{ij} = 1$, the triple (x_j, z_j, y_j) or its projection on the estimated frontier belong to a technology characterized by constant returns to scale (*crs*-technology). When $\hat{S}_{ij} < 1$, the observation or its projection is not on a vertex or a side where the technology is *crs*. When returns to scale are not constant, this approach can be extended to determine the exact nature of the returns to scale, as we will see below.

Obviously, the scale ratio is computed using an estimated frontier, not the true one. Suppose that for an observation, say (x_j, z_j, y_j) , we obtain $\hat{S}_{ij} < 1$, then without a formal statistical test procedure it is impossible to identify if this is a true scale inefficiency or the result of sampling variations. That is, since the scale ratio is calculated based on an estimated frontier, a score strictly smaller than one can be attributed to sampling variations. Consequently, we have to test the hypothesis that $S_{1j}(x_j, z_j, y_j) = 1$ using the statistic $\hat{S}_{ij}(x_j, z_j, y_j)$ to determine if the DMU is really scale inefficient.

The test procedure is in two steps. The first step consists in testing the null that a given triple (x, z, y) is scale efficient, i.e. its technology is of the *crs* type. The alternative hypothesis has to be less restrictive. One natural hypothesis is that the triple (x, z, y) is characterized by a *vrs*-technology. Then we have:

$$\begin{aligned} \text{Test \#1: } H_0 &: S_1(x, z, y) = 1 \\ H_A &: S_1(x, z, y) < 1 \end{aligned}$$

If the hypothesis H_0 is rejected, we still have to identify whether the returns to scale are increasing or decreasing. This will work if we can find a “new” null hypothesis that is less restrictive than the one in the first test. One way of doing this is to suppose that under the null hypothesis the triple (x, z, y) is subject to decreasing returns to scale (*drs*). Then, the alternative hypothesis would be that the triple (x, z, y) is subject to an increasing return to scale (*irs*) technology, since constant returns to scale have already been rejected in the first test. Thus, the second test is:

$$\begin{aligned} \text{Test \#2: } H_0' &: S_2(x, z, y) = 1 \\ H_A' &: S_2(x, z, y) < 1 \end{aligned}$$

The test statistics for the first and second test are $\hat{S}_1(x, z, y) = \hat{\theta}^{crs}(x, z, y) / \hat{\theta}^{vrs}(x, z, y)$ and $\hat{S}_2(x, z, y) = \hat{\theta}^{nirs}(x, z, y) / \hat{\theta}^{vrs}(x, z, y)$, respectively. Given a critical value $c_\alpha > 0$ for a given

test size, α , the decision rules are such that we reject H_0 each time $(1 - \hat{S}_1(x, z, y)) \geq c_\alpha$ in the first case and each time $(1 - \hat{S}_2(x, z, y)) \geq c_\alpha$ in the second case.

This test procedure is applied to each observation (x_j, z_j, y_j) in Ψ_0 . Consequently, we have at least n identical hypothesis testing procedure to run, leading to n decisions concerning the returns to scale for each of the n individual organizations represented in Ψ_0 .

To apply these tests we need to either choose the critical value c_α for a given test size or alternatively we can estimate the p -value for each test. In both cases, it is necessary to have the distribution of the test statistics. This is the problem discussed in the next section.

5. Bootstrapped test statistics

To determine the critical values and the decision rules, we use the approach developed in Essid, *et al.* (2010b), which is based on Simar and Wilson (2002). This approach allows us to find an approximation of the sampling distribution of the estimators of both scale efficiency ratios, \hat{S}_1 and \hat{S}_2 . This approximation rests on the bootstrap method that consists in identically replicating the empirical DGP many times and study the behavior this set of bootstrapped estimates. To implement the procedure, we first generate, from the original sample Ψ_0 , B pseudo-samples: Ψ_b^* , $b = 1, \dots, B$. Then, the original estimation method (DEA in our case) is applied to each pseudo-samples to obtain the bootstrap estimator of the test statistic \hat{S}_{1b}^* for \hat{S}_1 (and \hat{S}_{2b}^* for \hat{S}_2). This procedure allows us to estimate the empirical distribution of $(\hat{S}_1^* - \hat{S}_1)$ (and of course $(\hat{S}_2^* - \hat{S}_2)$), used to approximate the unknown distribution of the statistic $(\hat{S}_1 - S_1)$ (and $(\hat{S}_2 - S_2)$), under the null hypothesis.

The bootstrap relies on the principle that the pseudo-samples Ψ_b^* are generated so that:

$$(\hat{S}_1 - 1) \Big| \mathfrak{S}^{crs} \overset{approx}{\sim} (\hat{S}_1^* - \hat{S}_1) \Big| \hat{\mathfrak{S}}^{crs} \text{ for Test \# 1} \quad (10)$$

and

$$\left(\hat{S}_2 - 1\right) \Big| \mathfrak{S}^{nirs} \stackrel{approx}{\sim} \left(\hat{S}_2^* - \hat{S}_2\right) \Big| \hat{\mathfrak{S}}^{nirs} \text{ for Test \# 2,} \quad (11)$$

where $\hat{\mathfrak{S}}$ is a consistent estimator of the DGP \mathfrak{S} under the hull hypothesis, i.e. *crs* in the first case and *nirs* in the second case.

To generate the pseudo-efficiencies $\hat{\theta}^{*crs}$, $\hat{\theta}^{*vrs}$ and $\hat{\theta}^{*nirs}$, and then the statistics \hat{S}_1^* and \hat{S}_2^* , we first use a homogenous bootstrap methodology developed by Simar and Wilson (1998). This procedure rests on the assumption that the efficiency structure is homogenous. That is, the efficiency score θ is independent of (η, z, y) : $f(\theta|\eta, z, y) = f(\theta)$. A consistent estimator of f , obtained using a kernel estimator and corrected by Schuster's (1985) and Silverman (1985) is defined as follows:

$$\hat{f}^c(t) = \begin{cases} 2\hat{g}(t) & \text{if } t \leq 1 \\ 0 & \text{otherwise} \end{cases}, \text{ where } \hat{g}(t) = \frac{1}{2nh} \sum_{j=1}^{j=n} \left[\phi\left(\frac{t - \hat{\theta}_j}{h}\right) + \phi\left(\frac{t - 2 + \hat{\theta}_j}{h}\right) \right]. \quad (12)$$

We use a normal Gaussian kernel, denoted ϕ , and the bandwidth, h , is set following the normal reference rule (Silverman (1986)).

The robustness of the testing procedure with respect to the simulations is checked by relaxing the homogeneity assumption in a second simulation experiment. We run a smooth heterogeneous bootstrap simulation, based on a methodology developed in Simar and Wilson (2000). Then the test procedure is re-run with the new bootstrapped dataset. In this case, the density f is estimated by:¹

$$\hat{f}^c(u) = \begin{cases} 2\hat{f}(u) & \text{if } u \in \Omega \\ 0 & \text{otherwise} \end{cases}, \quad (13)$$

where $\hat{f}(u) = \frac{1}{2nh^{m_1+m_2+s}} \sum_{j=1}^{j=n} \left[K_1\left(\frac{u - u_j}{h}\right) + K_2\left(\frac{u - u_{Rj}}{h}\right) \right]$, $K_l(\cdot)$ is the probability density of a normal vector with zero mean and variance-covariance matrix $\hat{\Sigma}_l$, $l=1,2$, $u_j = (\theta_j, \eta_j, z_j, y_j)$,

¹ Both simulation procedures are presented in Essid *et al.* (2010a).

and $u_{Rj} = (2 - \theta_j, \eta_j, z_j, y_j)$. The bandwidth parameter is set according to Silverman's normal rule.

The pseudo-scores are generated in three steps. We begin by presenting the procedure for the first test:

Step 1: Generate $\hat{\theta}^*$ from $\hat{\theta}^{crs}$.

Step 2: Generate the pseudo-sample $\Psi^* = \{(x^*, z^*, y^*)\}$ using the estimator \hat{f}^c of f given by equation (12), in the case of the homogenous bootstrap and using equation (13) in the case of the heterogeneous bootstrap.

Step 3: Compute $\hat{\theta}^{*crs}$ and $\hat{\theta}^{*vrs}$ using the same pseudo sample Ψ^* for both measures, insuring that the pseudo-score $\hat{S}_1^* = \hat{\theta}^{*crs} / \hat{\theta}^{*vrs}$ has an admissible value, i.e. $\hat{S}_1^* = \hat{\theta}^{*crs} / \hat{\theta}^{*vrs} \leq 1$.²

This three step procedure is then replicated B times to obtain the statistic $\hat{S}_{1j} = \hat{\theta}^{crs}(x_j, z_j, y_j) / \hat{\theta}^{vrs}(x_j, z_j, y_j)$ for each triple (x_j, z_j, y_j) in the dataset Ψ_0 . Then we obtain B pseudo-values $\hat{S}_{1jb}^* = \hat{\theta}_b^{*crs}(x_j, z_j, y_j) / \hat{\theta}_b^{*vrs}(x_j, z_j, y_j)$, $b = 1, \dots, B$ for all $j = 1, \dots, n$.

To obtain the pseudo-scores $\hat{S}_2^* = \hat{\theta}^{*nirs} / \hat{\theta}^{*vrs}$, we only need to bootstrap $\hat{\theta}^{nirs}$ instead of $\hat{\theta}^{crs}$ in Step 1 above to recover the pseudo-sample Ψ^* used to compute the pseudo-efficiencies. Then, the test statistic $\hat{S}_{2j} = \hat{\theta}^{nirs}(x_j, z_j, y_j) / \hat{\theta}^{vrs}(x_j, z_j, y_j)$ is replicated B times, i.e. $\hat{S}_{2jb}^* = \hat{\theta}_b^{*nirs}(x_j, z_j, y_j) / \hat{\theta}_b^{*vrs}(x_j, z_j, y_j)$, $b = 1, \dots, B$, for each triple in Ψ_0 .

To obtain the critical values and infer the decision rules of both tests, we use the bootstrap analog of the true size of the test, denoted α (i.e. the probability of Type 1 error). For the first and second tests they are respectively defined as $\Pr(\hat{S}_1^* \leq \hat{S}_1 - c_\alpha^* | H_0, \hat{\mathcal{S}}^{crs}) = \alpha$ and $\Pr(\hat{S}_2^* \leq \hat{S}_2 - c_\alpha^* | H_0, \hat{\mathcal{S}}^{nirs}) = \alpha$, where c_α^* is an approximation of the critical value c_α .

² This at this stage that we make sure that the DGP is consistent with the hypothesis tested. With this procedure, the test statistic is always less than or equal to one because both efficiency measures are calculated from data coming from the same DGP.

To obtain these critical values, we start by sorting in ascending order the pseudo-scores $(\hat{S}_{1b}^* - \hat{S}_1)$ (or $(\hat{S}_{2b}^* - \hat{S}_2)$) for $b = 1, \dots, B$. Then, we eliminate $(1 - \alpha) \times 100$ percent of the values at the right end of the ordered sequence of bootstrapped statistics. The critical value $-c_\alpha^*$ is then the right end of this truncated sequence. Thus, the decision rule is the following: reject the null hypothesis if, for a given size α , $\hat{S}_1 \leq 1 - c_\alpha^*$ for the first test and $\hat{S}_2 \leq 1 - c_\alpha^*$ for the second.

Hypothesis tests can be more convincingly used by giving the marginal probability of rejections, or p -value. The bootstrap analog of the p -value for H_0 is $\hat{p} = \Pr(\hat{S}_1^* \leq \hat{S}_1 | H_0, \hat{\mathcal{S}}^{crs})$ for the first test and $\hat{p} = \Pr(\hat{S}_2^* \leq \hat{S}_2 | H_0, \hat{\mathcal{S}}^{nirs})$ for the second. For each test, we reject null hypothesis when $\hat{p} \leq \alpha$, for a given size α .

6. Application to Tunisian high schools

The report of the World Bank (2008) on the state of education on the Middle East and North Africa (MENA) concluded that the education system does not satisfy the needs of the society and does not meet the growing ambitions of the people. It is said that the failure to achieve positive results rests mostly on a misallocation of the resources devoted to the education system and to schools operating inefficiently. In the case of Tunisia, Essid *et al.* (2010a) show that high schools would be able to deliver the same amount of services with 12% less resources on average. Other than the internal working of institutions, the overall organization of the education system has been under scrutiny.

To improve the internal performance of the education system, the department of education has made a priority of reducing the size of the schools. Table 1 shows that the number of schools and the number of teachers have increased during the period 2001-2005 to keep up with the constant increase in enrolment. It is noteworthy that the number of schools has increased sufficiently rapidly that the average size of the schools, measured by the number of students, has decreased over the same period: from 965 in 2001/2002 it fell to 946 in 2002/2003 and to 911 in 2004/2005. That is, the average size of the Tunisian secondary education establishments has continuously decreased since 2001. This is consistent with the belief of the Tunisian department of education that smaller units are performing better. This clearly raises the question concerning

the scale characterization of schools and how the optimal size must be determined so that the schools operate at their optimal scale.

Table 1: Evolution of the number of schools and their average size

Academic Year	Number of Schools	Enrolment	Number of student-class	Number of Teachers	Average School Size
01/02	1,065	1,027,812	31,146	49,331	965
02/03	1,117	1,057,233	32,300	51,738	946
03/04	1,161	1,076,238	33,103	55,717	927
04/05	1,191	1,084,878	33,811	59,132	911

Source: « Bureau des études, de la planification et de la programmation », Department of Education, Tunisia

Here we would like to characterize the returns to scale and test whether the government is right about the optimal size of schools. In order to do this, we apply our bootstrap methodology to a sample of 332 Tunisian high schools. (This sample contains roughly 28% of the schools for the reference year 2003/2004.). The *Bureau des études, de la planification et de la programmation* of the Tunisian department of education has provided data for each school for the academic year 2003/2004. The data include total enrolment, the number of students graduating (successfully completing the *baccalauréat*), the number of teachers and non-teaching staffs, the number of students in residence (accommodation provided by the institution), the number of general teaching and specialized classrooms, and the operating budget. We have used price indices from the Tunisian National Institute of Statistics (consumer price index for 2004). Using these data, we have constructed our output measures and the quantities of quasi-fixed and variable inputs for each high school.³

Tunisian high schools supply two types of services: education and residence (full-board accommodation). We use the number of beds (BEDS) and the number of meals served (MEALS) to measure the residence services. The education services are measured using total enrolment (STUDENTS) and the *baccalauréat*⁴ results (RBAC) which corresponds to the number of successful students per school.

To produce “students”, high schools use variable and quasi-fixed inputs. The variable inputs are the total number of teachers (TEACHERS), the number of administrative personnel (ADM), the number of blue collars, essentially maintenance and cleaning staff (BLUECOL), and

³ For details on the data construction the reader is referred to Essid *et al.* (2010a).

⁴ The *baccalauréat* is the diploma obtained at the end of high school; it is the equivalent of a high school diploma

finally from the operating budget and the consumer price index, we construct a proxy for food and material supplies (F&MAT). The quasi-fixed factors are given by the number of general classrooms (GROOM) and the number of specialized classrooms (SROOM). Descriptive statistics of the data are given in Table 2.

Table 2: Descriptive Statistics

	Variable	Mean	Standard error	Minimum	Maximum
Outputs	STUDENTS	1292,25	461,64	346	2769
	BEDS	123,65	188,31	0	931
	MEALS	173,35	220,00	0	931
	RBAC	180,66	94,05	18	526
Variable inputs	TEACHERS	72,53	24,32	26	152
	ADM	9,53	4,68	1	28
	BLUECOL	14,11	7,63	3	48
	F&MAT	641,95	377,63	147,08	1983,86
Quasi-fixed inputs	GROOM	26,18	8,77	11	59
	SROOM	10,58	4,37	3	31

Table 3 and 4 contains a summary of the scale efficiency measurement results.⁵ It includes the standard DEA results and the test of hypothesis using the homogenous and heterogeneous bootstrap simulations of our model. The number of bootstrap simulation is equal to 2,000 ($B=2,000$).

Table 3 presents the results with the schools grouped by size (measured by the number of students). Small schools have less than one thousand students (97 schools), the middle group (129 schools) includes schools with a number of students between 1,000 and 1,500 and the last group includes the large schools, that is, the schools with more than 1,500 students (106 schools).⁶ The total number of schools is 332. Except for the first line, each cell contains a triple. The three entries are respectively the number of schools that exhibits increasing returns to scale, constant returns to scale and decreasing returns to scale.

⁵ Full results are in an Appendix available from the authors upon request.

⁶ There is no formal definition of “small”, “medium” and “large” institutions in the literature. The corresponding sizes change considerably from one study to the other (see Stiefel *et al.* (2009) for a complete survey of the problem). We have kept in this paper a breakdown of the schools sizes such that the distribution is relatively homogenous between groups.

Table 3: Results of the test for the homogenous and the heterogeneous bootstrap

	School Size			Total
	Small	Medium	Large	
Number of schools	97	129	106	332
Scale Efficiency: Number of schools per RTS type (<i>irs</i> , <i>crs</i> , <i>drs</i>)	(85, 11, 1)	(86, 23, 20)	(26, 13, 67)	(197, 47, 88)
Scale Efficiency Test:				
1) Homogenous Bootstrap:				
$\alpha = 1\%$	(20, 77, 0)	(6, 123, 0)	(0, 106, 0)	(26, 306, 0)
$\alpha = 5\%$	(35, 62, 0)	(14, 115, 0)	(0, 105, 1)	(49, 282, 1)
$\alpha = 10\%$	(46, 51, 0)	(20, 109, 0)	(0, 101, 5)	(66, 261, 5)
2) Heterogeneous Bootstrap				
$\alpha = 1\%$	(33, 64, 0)	(0, 129, 0)	(0, 106, 0)	(33, 299, 0)
$\alpha = 5\%$	(10, 87, 0)	(0, 129, 0)	(0, 106, 0)	(10, 322, 0)
$\alpha = 10\%$	(19, 78, 0)	(0, 129, 0)	(0, 106, 0)	(19, 313, 0)

We first present the traditional scale efficiency scores proposed by Färe *et al.* (1994). The results are grouped by school sizes and each triple contains the number of schools belonging to each category of returns to scale. Clearly, most DMUs are not scale efficient (more than 85%). More than half of the schools exhibit increasing returns to scale, this result being concentrated in the small and medium size schools (85 small and 86 medium size high school). The majority of the large schools (67 out of 106) exhibit decreasing returns to scale. There are fourteen percent of the high schools that exhibit constant returns to scale. These schools are deemed scale efficient. Half of these schools are medium size institutions. The other half is spread evenly between the small and large schools. Clearly, the first analysis using the standard approach does not provide any indication on the optimal size for a school. Therefore, we have to look for an answer somewhere else in order to get guidance in choosing the optimal size. A potential direction to look for is suggested by the fact that these results on the optimal size are based on the raw scores and do not take into account potential sampling variations.

Such a traditional presentation of the results does not include the possibility that the scale efficiency scores might be estimators of a parameter. As explained above, the scores are not measured with respect to the true frontier, but with respect to an estimated frontier. Consequently, the scale efficiency scores are themselves estimates of the true scores, thus subsumed to statistical regularities; hence the results reported above must be accompanied by an estimate of their

1
2
3
4
5
6 precision. In other words, reported cases of scale inefficiencies in the traditional approach might
7 be the result of sampling variations.
8
9

10 The bootstrap general simulations results are displayed in last rows of Table 3, while
11 Table 4 presents some specific cases. The simulations results are always reported for three
12 different test sizes: $\alpha = 1\%$, $\alpha = 5\%$ and $\alpha = 10\%$. The application of the test procedure we
13 developed in the previous sections confirms the intuition that sampling variations play a
14 significant role in the estimation of scale efficiencies. The bottom part of Table 3 confirms that
15 the vast majority of the schools operate under constant returns to scale. That is, more than 78% of
16 the institutions are now deemed scale efficient at all test sizes for both simulation methods, in
17 strong contrast with what we found initially. In other words, the scale inefficiencies identified by
18 the brute scores are spoiled by some noise and most of the decision units are in fact scale
19 efficient. These results are not sensitive to the type of bootstrap procedure, homogenous or
20 heterogeneous, so even the simplest statistical procedure raises some concerns on the
21 interpretation of the straight scores of scale efficiencies without its statistical content. The results
22 in Table 3 allow us to show the existence of a significant relationship between high school sizes
23 and scale efficiency scores. The highest concentration of scale-efficient schools is found among
24 the medium or large size institutions. At test size $\alpha = 5\%$, out of the 282 schools deemed scale
25 efficient using the homogenous bootstrap method, 115 are medium size high schools and 105 are
26 large institutions. For the same test size, the results using the heterogeneous bootstrap simulations
27 lead us to a fairly similar conclusion (out of 322 scale efficient institutions we have 122 medium
28 and 106 large institutions). This also raises the point that when a school is small it is more likely
29 to be scale-inefficient. This casts some doubts on the policy promoted by the department of
30 education that organized a systematic reduction of the school sizes. In lights of our results, this
31 movement is not totally optimal, as in general smaller schools are more likely to have difficulties
32 to be scale efficient than medium and large size institutions.
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

50 Table 4 presents the result for specific units. Our approach has focused on unit specific
51 test procedure by contrast to Simar and Wilson (2002) that consider a global test. It compares
52 directly the simulation results obtained using the homogenous and heterogeneous bootstrap to the
53 standard results (ignoring the statistical content of the analysis). Some specific examples of
54 DMUs are presented in the Table and serve to illustrate some general trends.
55
56
57
58
59
60

[INSERT TABLE 4 HERE]

The homogenous bootstrap results confirm our intuition as it is impossible to reject the hypothesis of scale efficiency (*crs*) at a size equal to 5% for 282 high schools (more than 84% of the sample). Most of those schools are of either medium or large size (roughly 80% of the scale efficient schools). There are 49 high schools (more than 14% of the sample) that exhibit increasing returns to scale. Most of these institutions are small schools. That is, 35 high schools (70% of the group) have less than one thousand students. High school L1115, with an enrolment of 1801 students, is the only case of decreasing returns to scale (significant at 5%) and not surprisingly it is a large school. These results are robust to the size. At size $\alpha = 10\%$, we continue to observe a fairly large proportion of high schools, mostly medium and large institutions, that exhibits constant returns to scale (more than 78% of the sample). There is also a smaller proportion of the high schools (about 20%), mostly of a small size, that exhibit increasing returns to scale. There are only at most five high schools which can be deemed to operate under decreasing returns to scale and they are all large institutions.

In theory, the heterogeneous bootstrap simulations would provide results closer to the reality of the institutions because they account for the specificities of the production process. The cost for this is important as the procedure is quite involved. The comparison of its performance with the homogenous procedure is thus interesting. The last columns of Table 4 present specific cases for this procedure and must help to understand what is going on in both procedures. A striking point is that the results are very similar to those of the homogenous bootstrap, something we interpret as a confirmation of the robustness of the test procedures and their results. The number of scale efficient DMU increases to 90% for all three test sizes. In both simulations, most of the DMU exhibit the same type of returns to scale. The heterogeneous bootstrap tends to ascribe the scale efficient label to schools more often than the homogenous procedure, however. None of the high schools exhibit decreasing returns to scale at all three sizes for the heterogeneous bootstrap. One striking feature of these results is the relationship between the types of returns to scale and the enrolment, our size indicator of the schools. The heterogeneous bootstrap simulations are clear on that point: schools of medium and large sizes are all scale efficient. That is, they are deemed to have an optimal size at α equal to 1%, 5% and 10%. The

1
2
3
4
5
6 DMU that are not efficient, for which the null hypothesis is rejected at α equal to 1% in the first
7 test, are all schools with less than one thousand students, the smallest schools of our sample.
8
9

10 This result is in line with the conclusion we reached for the homogenous bootstrap
11 simulations. It seems that most medium and large schools are scale efficient, while their smaller
12 counterparts are not. The latter exhibit increasing returns to scale and from the inverse
13 relationship identified between returns to scale and size measured by the number of students, we
14 would be inclined to believe that these schools are too small. Clearly, these conclusions are not
15 good for the policy promoted by the Tunisian department of education. It seems that the route to
16 improve the efficiency of the education system in Tunisia does not go through a systematic
17 reduction of the size of the high schools. To the contrary, the smallest institutions tend to boast
18 bad results in terms of scale efficiency. Some small schools seem to sit on potential performance
19 gains that cannot be exploited due to their size.
20
21
22
23
24
25
26

27 **7. Conclusion**

28
29
30 A very entrenched claim by program designers and education system decision makers is that
31 small establishments are better than bigger one for the students (Schneider *et al.* (2007) for a
32 discussion of this point). Despite the pedagogical arguments to that effect, for economists this
33 problem is directly related to the scale efficiency. In other words, what is the school optimal size
34 so that all resources devoted to the institution are optimally used to produce well educated
35 students? To respond to this question and orient correctly the education policy we need a formal
36 way to determine the optimal size of given educational institutions.
37
38
39
40
41
42

43 In this article we have developed a procedure to test non parametric statistical hypothesis
44 concerning the scale efficiency of organization, readily implementable to schools. Because there
45 is no asymptotic distribution for the test statistic under the null hypothesis, we used a smooth
46 bootstrap methodology to approximate it. This approach allows us to estimate the p -values of the
47 test and to determine a decision rule to accept or reject the null hypothesis. The results of our
48 application of this methodology to the Tunisian secondary schools show on the one hand that
49 scale efficiencies are strongly sensitive to sampling variation and on the other hand that the claim
50 made by the decision makers in the Department of Education in Tunisia that the country should
51 aimed at reducing as much as possible the size of the schools is not verified. Smaller schools are
52 not better and in fact we have shown that they are probably the only one that are systematically
53
54
55
56
57
58
59
60

not of the right size as they are exhibiting increasing returns to scale. It is in medium and large institutions that we can find high schools fully benefiting from scale efficiencies.

We have ignored the socio-economic characteristics of students and the localization of the schools, however. This may eventually force us to qualify our results as rural schools may not be able to increase the number of students because of weak population density for instance. Nonetheless, we do not expect these considerations to change radically the picture we drew of the school efficiency in Tunisia.

References

- Banker, R.D. (1984), "Estimating Most Productive Scale Size Using Data Envelopment Analysis", *European Journal of Operational Research* 17, 35-44.
- Banker, R.D. and Morey, R.C. (1986), "Efficiency Analysis for Exogenously Fixed Inputs and Outputs", *Operations Research* 34, 513-521.
- Banker, R.D. and Thrall, R.M. (1992), "Estimation of Returns to Scale Using Data Envelopment Analysis", *European Journal of Operational Research* 62, 74-84.
- Banker, R.D., Chang, H. and Cooper, W.W. (1994), "Equivalence of Alternative Methods for Returns-to-Scale Estimation in Data Envelopment Analysis", *Unpublished working paper*.
- Banker, R.D., Charnes, A. and Cooper, W.W. (1984), "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis", *Management Science* 30, 1078-1092.
- Chabotar, K. J. (1989), "Measuring the Costs of Magnet Schools", *Economics of Education Review* 8, 169-183.
- Colegrave, A. D. and Giles, M. J. (2005), "School Cost Functions: A Meta Regression Analysis", Center for Labor Market Research, University of Western Australia Crawley, WA, paper.
- Coladarci, T. and Cobb, C. D. (1996), "Extracurricular Participation, School Size, and Achievement and Self-esteem among High School Students: A National Look", *Journal of Research in Rural Education* 12, 92-103.
- Essid, H., Ouellette, P. and Vigeant, S. (2010a), "Measuring Efficiency of Tunisian Schools in the Presence of Quasi-fixed Inputs: A Bootstrap DEA Approach", *Economics of Education Review* 29, 589-596.
- Essid, H., Ouellette, P. and Vigeant, S. (2010b), "Scale Efficiency Tests and DEA Estimation: A Bootstrap Approach ", mimeo, USTL.

- 1
2
3
4
5
6 Essid, H., Ouellette, P. and Vigeant, S. (2010c), "A Note on the Convergence of Nonparametric
7 DEA Estimators for Production Efficiency Scores with Quasi-Fixed Inputs", mimeo,
8 Institut des HEC de Carthage.
9
10
11 Farrell, M.J. (1957), "The Measurement of Productive Efficiency", *J.Roy.Stat.Soc. A* 120, 253-
12 290.
13
14
15 Foreman-Peck, J. and Foreman-Peck, L. (2006), "Should Schools Be Smaller ? The Size-
16 Performance Relationship for Welsh Schools", *Economics of Education Review* 25, 157-
17 171.
18
19
20 Fox, W. F. (1981), "Reviewing Economies of Size in Education", *Journal of Education Finance*
21 6, 273-296.
22
23
24 Färe, R. (1988), "Fundamentals of Production Theory", *Lecture Notes in Economics and*
25 *Mathematical Systems*, Vol. 311, Springer-Verlag, Berlin Heidelberg.
26
27 Färe, R. and Grosskopf, S., (1994), "Estimation of Returns to Scale Using Data Envelopment
28 Analysis: A Comment", *European Journal of Operational Research* 79, 379-382.
29
30 Färe, R. and Grosskopf, S. (1985), "A Nonparametric Cost Approach to Scale Efficiency",
31 *Scandinavian Journal of Economics* 87, 594-604.
32
33
34 Färe, R., Grosskopf, S. and Lovell C.A.K. (1994), "Production Frontiers", *Cambridge University*
35 *Press*, Cambridge.
36
37 Kneip, A., Park, B.U. and Simar, L. (1998), "A Note on the Convergence of Nonparametric DEA
38 Estimators for Production Efficiency Scores", *Econometric Theory* 14, 783-793.
39
40
41 Kumar, R. C. (1983), "Economies of Scale in School Operation: Evidence from Canada", *Applied*
42 *Economics* 15, 323-340.
43
44
45 Kuziemko, I. (2006), "Using Shocks to School Enrollment to Estimate the Effect of School Size
46 on Student Achievement", *Economics of Education Review* 25, 63-75.
47
48
49 Leithwood K. and Jantzi D. (2009), "A Review of Empirical Evidence about School Size Effects:
50 A Policy Perspective", *Review of Educational Research* 79, 464-490.
51
52
53 McGuire, K. (1989), "School Size: The Continuing Controversy" *Education and Urban Society*
54 21, 164-174.
55
56
57 Monk, D. H. (1987), "Secondary School Size and Curriculum Comprehensiveness", *Economics*
58 *of Education Review* 6, 137-150.
59
60

- 1
2
3
4
5
6 Morris, H. J. (1964), *Relationship of School Size to Per Pupil Expenditures in Secondary Schools*
7 *in the Southern Region*, Unpublished doctoral dissertation, George Peabody College for
8 Teachers, Nashville, TN.
9
10
11 Park, B. U., Jeong, S. –O. and Simar, L. (2010), "Asymptotic Distribution of Conical-Hull
12 estimators of Directional Edges", *The Annals of Statistics* 38, No. 3, 1320-1340.
13
14
15 Schneider, B., Wyse, A. E., and Keesler, V. (2007), "Is Small Really Better ? Testing Some
16 Assumptions about High School Size", In L. Tom & F.M. Hess (Eds.), *Brookings Papers*
17 *on Education Policy 2006/2007*, Washington, DC: The Brookings Institution.
18
19
20 Schuster, E.F. (1985), "Incorporating Support Constraints into Nonparametric Estimators of
21 Densities", *Communications in Statistics-Theory and methods* 14, 1123-1136.
22
23
24 Silverman, B.W. (1986), "Density Estimation for Statistics and Data Analysis", *Chapman and*
25 *Hall*.
26
27 Simar, L. and Wilson, P.W. (2002), "Non-Parametric Tests of Returns to Scale", *European*
28 *Journal of Operational Research* 139, 115-132.
29
30 Simar, L. and Wilson, P.W. (2000), "A General Methodology for Bootstrapping in non-
31 parametric frontier models", *Journal of Applied Statistics* 27, 779-802.
32
33
34 Simar, L. and Wilson, P.W. (1998), "Sensitivity Analysis of Efficiency Scores: How to Bootstrap
35 in Nonparametric Frontier Models", *Management Science* 44, 49-61.
36
37 Stiefel, L., Schwartz, A.E., Iatarola, P. and Chellman, C.C. (2009), "Mission Matters: The Cost of
38 Small High Schools Revisited", *Economics of Education Review* 28, 585-599.
39
40
41 Turner, C. C., and Thrasher, J. M. (1970), "School Size Does Make a Difference", San Diego,
42 CA: Institute for Educational Management (ERIC Document Reproduction Service No.
43 ED 043946).
44
45
46 Word Bank, (2008), *The Road Not Traveled Education Reform in the Middle East and North*
47 *Africa. The World Bank*, Washington, D.C.
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 4: Homogenous and the heterogeneous bootstrap test results: some specific cases

DMU	Enrolment	\hat{S}_{1j}	\hat{S}_{2j}	Initial Scale efficiency	Homogenous bootstrap simulation results						Heterogeneous bootstrap simulation results					
					\hat{p}_j (Test 1)	\hat{p}_j (Test 2)	Scale efficiency test			\hat{p}_j (Test 1)	\hat{p}_j (Test 2)	Scale efficiency test				
							$\alpha=0,01$	$\alpha=0,05$	$\alpha=0,1$			$\alpha=0,01$	$\alpha=0,05$	$\alpha=0,1$		
L111	1517	0,9986	0,9986	<i>Irs</i>	0,6775	0,2855	<i>crs</i>	<i>crs</i>	<i>crs</i>	0,8305	0,5325	<i>crs</i>	<i>crs</i>	<i>crs</i>		
L1111	1205	0,9929	0,9929	<i>Irs</i>	0,4950	0,4490	<i>crs</i>	<i>crs</i>	<i>crs</i>	0,6195	0,4725	<i>crs</i>	<i>crs</i>	<i>crs</i>		
L1112	1910	0,9696	1,0000	<i>drs</i>	0,3730	0,9555	<i>crs</i>	<i>crs</i>	<i>crs</i>	0,5790	0,9450	<i>crs</i>	<i>crs</i>	<i>crs</i>		
L1113	1875	1,0000	1,0000	<i>drs</i>	0,1995	0,9695	<i>crs</i>	<i>crs</i>	<i>crs</i>	0,9075	0,9405	<i>crs</i>	<i>crs</i>	<i>crs</i>		
L1115	1801	0,9294	1,0000	<i>drs</i>	0,0395	0,9045	<i>crs</i>	<i>drs</i>	<i>drs</i>	0,4040	0,9480	<i>crs</i>	<i>crs</i>	<i>crs</i>		
L1126	1746	1,0000	1,0000	<i>crs</i>	0,9570	0,9690	<i>crs</i>	<i>crs</i>	<i>crs</i>	0,9850	0,9515	<i>crs</i>	<i>crs</i>	<i>crs</i>		
L1235	632	0,9630	0,9630	<i>irs</i>	0,5420	0,5420	<i>crs</i>	<i>crs</i>	<i>crs</i>	0,3620	0,3280	<i>crs</i>	<i>crs</i>	<i>crs</i>		
L1545	1404	1,0000	1,0000	<i>crs</i>	0,9360	0,9925	<i>crs</i>	<i>crs</i>	<i>crs</i>	0,9845	0,9915	<i>crs</i>	<i>crs</i>	<i>crs</i>		
L1464	669	0,8910	0,8910	<i>irs</i>	0,0005	0,0005	<i>irs</i>	<i>irs</i>	<i>irs</i>	0,1480	0,1455	<i>crs</i>	<i>crs</i>	<i>crs</i>		
L3183	615	0,8549	0,8549	<i>irs</i>	0,0090	0,0090	<i>irs</i>	<i>irs</i>	<i>irs</i>	0,1085	0,1060	<i>crs</i>	<i>crs</i>	<i>crs</i>		
L41120	1564	0,9874	1,0000	<i>drs</i>	0,0955	0,9740	<i>crs</i>	<i>crs</i>	<i>drs</i>	0,7040	0,9495	<i>crs</i>	<i>crs</i>	<i>crs</i>		
L43134	962	0,9446	0,9446	<i>irs</i>	0,0115	0,0115	<i>crs</i>	<i>irs</i>	<i>irs</i>	0,3475	0,3020	<i>crs</i>	<i>crs</i>	<i>crs</i>		
L43143	1689	0,9608	1,0000	<i>drs</i>	0,0540	0,8920	<i>crs</i>	<i>crs</i>	<i>drs</i>	0,5105	0,9500	<i>crs</i>	<i>crs</i>	<i>crs</i>		
L52168	1620	0,9844	1,0000	<i>irs</i>	0,0850	0,1995	<i>crs</i>	<i>crs</i>	<i>drs</i>	0,6710	0,6060	<i>crs</i>	<i>crs</i>	<i>crs</i>		
L84305	1910	0,9254	1,0000	<i>drs</i>	0,0695	0,8880	<i>crs</i>	<i>crs</i>	<i>drs</i>	0,3940	0,9405	<i>crs</i>	<i>crs</i>	<i>crs</i>		