Efficiency in Education Production among PISA Countries, with Emphasis on Transitioning Economies

Paul W. Wilson*
Department of Economics
University of Texas
1 University Station, C3100
Austin, Texas 78712 USA
wilson@eco.utexas.edu

June 2005

*Financial support from the Internation Bank for Reconstruction and Development (World Bank) is gratefully acknowledged. I have benefited from comments by Emanuala di Gropello, Sergei Soares, and Joe Shapiro. Any errors are solely my responsibility.

1 Introduction

The Organization for Economic Co-operation and Development (OECD)'s Programme for International Student Assessment (PISA) is an internationally standardized assessment that was jointly developed by participating countries and administered to 15-year-old students. PISA assesses how far students near the end of compulsory education have acquired some of the knowledge and skills that are essential for full participation in society. In all cycles, the domains of reading, mathematical and scientific literacy are covered not merely in terms of mastery of the school curriculum, but in terms of important knowledge and skills needed in adult life.

PISA provides a rare opportunity for cross-country comparisons. This study uses data from year 2000, the first assessment conducted by PISA, to answer several questions:

1. How efficient are schools in transitioning countries in the former Soviet Union relative to schools in other regions?

2. Which transitioning countries are most efficient in providing educational services?

3. How wide is the variation in efficiency within transitioning countries?

4. Are there combinations of inputs that appear particularly productive?

5. What improvements in outcomes might be possible? Alternatively, what decreases in inputs might be possible without reducing output levels?

Examination of efficiency in provision of educational services is problematic for several reasons. There is no single, encompassing measure of output. Further, neither prices of inputs nor of outputs are observed. Not only are prices unobserved, but it is difficult to conceive what these prices might be given the lack of market mechanisms that would determine economically meaningful prices. These factors weigh in favor of modeling the production process in terms of an input/output mapping and employing non-parametric efficiency estimators.

The remainder of this paper unfolds as follows. Rigorous discussions of the statistical model and properties of the estimators are given in Sections 2 and 3. Section 4 discusses the data used for estimation, and empirical results are analyzed in Section 5. The last section gives some final remarks, including caveats and possible directions for future research.

2 A Statistical Model of Production

Before anything can be estimated, a statistical model must be defined; otherwise, it is not possible to know what is estimated. Denote the production possibilities set by

$$\mathcal{P} = \{(\boldsymbol{x}, \boldsymbol{y}) | \boldsymbol{x} \text{ can produce } \boldsymbol{y}\}, \tag{2.1}$$

where $\boldsymbol{x} \in \mathbb{R}_+^p$ and $\boldsymbol{y} \in \mathbb{R}_+^q$ denote vectors of inputs and outputs, respectively. The production possibilities set can be described in terms of its sections

$$\mathcal{X}(\boldsymbol{y}) = \{\boldsymbol{x} \in \mathbb{R}_+^p \mid (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}\}, \tag{2.2}$$

and

$$\mathcal{Y}(\boldsymbol{x}) = \{\boldsymbol{y} \in \mathbb{R}_+^q \mid (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}\}, \tag{2.3}$$

or input requirement sets and output correspondence sets, respectively.

The economic assumptions of Shephard (1970) and Färe (1988) are typical in studies of productive efficiency, and are adopted here. These include:

Assumption 2.1. *$\mathcal{P}$ is convex, $\mathcal{X}(\boldsymbol{y})$ is convex and closed for all $y \in \mathbb{R}_+^q$, and $\mathcal{Y}(\boldsymbol{x})$ is convex, bounded, and closed for all $\boldsymbol{x} \in \mathbb{R}_+^p$.*

Assumption 2.2. *All production requires the use of some inputs, i.e., $(\boldsymbol{x}, \boldsymbol{y}) \notin \mathcal{P}$ if $\boldsymbol{y} \geq 0, \ \boldsymbol{x} = 0$.*

Assumption 2.3. *Both inputs and outputs are strongly disposable, i.e., if $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}$ then $\widetilde{\boldsymbol{x}} \geq \boldsymbol{x} \Rightarrow (\widetilde{\boldsymbol{x}}, \boldsymbol{y}) \in \mathcal{P}$ and $\widetilde{\boldsymbol{y}} \leq \boldsymbol{y} \Rightarrow (\boldsymbol{x}, \widetilde{\boldsymbol{y}}) \in \mathcal{P}$.*

Assumption 2.1 is needed for consistency of DEA estimators, but not for consistency of FDH estimators. Assumption 2.2 merely prohibits the possibility of a “free lunch”; in other words, if any output is produced, some strictly positive quantity of at least one input must be used to do so.

The upper boundary of $\mathcal{P}$, denoted $\mathcal{P}^\partial$, is sometimes referred to as the *technology* or the *production frontier,* and is given by the intersection of $\mathcal{P}$ and the closure of its compliment. Assumption 2.3 assumes that inputs or outputs can be ignored at no consequence to feasible production; in other words, Assumption 2.3 is equivalent to an assumption of monotonicity for $\mathcal{P}^\partial$. Similarly, the closure of the compliment of $\mathcal{X}(\boldsymbol{y})$—denoted $\mathcal{X}^\partial(\boldsymbol{y})$—represents an isoquant. The closure of the compliment of $\mathcal{Y}(\boldsymbol{x})$, denoted $\mathcal{Y}^\partial(\boldsymbol{x})$, gives an iso-output or product transformation curve.

Different measures of technical inefficiency are sometimes used. The Shephard (1970) input distance function measures distance from an arbitrary point $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_+^{p+q}$ to $\mathcal{P}^\partial$ in a direction orthogonal to $\boldsymbol{y}$, and is defined by

$$\mathcal{I}(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}) \equiv \sup \left\{\theta > 0 \mid (\theta^{-1}\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}\right\}. \tag{2.4}$$

Similarly, the Shephard (1970) output distance function measures distance from an arbitrary point $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_+^{p+q}$ to $\mathcal{P}^\partial$ in a direction orthogonal to $\boldsymbol{x}$, and is defined by

$$\mathcal{O}(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}) \equiv \inf \left\{\theta > 0 \mid (\boldsymbol{x}, \theta^{-1}\boldsymbol{y}) \in \mathcal{P}\right\}. \tag{2.5}$$

For $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}$, $\mathcal{I}(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}) \geq 1$ and $\mathcal{O}(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}) \leq 1$. In addition, $\mathcal{I}(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P})$ gives the maximum, feasible proportionate reduction in inputs (holding output levels fixed), while $\mathcal{O}(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P})$ gives the maximum, feasible proportionate expansion in outputs (holding input levels fixed) for the point $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}$. For a point $(\boldsymbol{x}, \boldsymbol{y})$ in the interior of the production set, *i.e.* $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}$ and $(\boldsymbol{x}, \boldsymbol{y}) \notin \mathcal{P}^\partial$, efficient projections onto the frontier $\mathcal{P}^\partial$ are given by $(\boldsymbol{x}\mathcal{I}(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P})^{-1},\ \boldsymbol{y})$ in the input direction and $(\boldsymbol{x},\ \boldsymbol{y}\mathcal{O}(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P})^{-1},\ \boldsymbol{y})$ in the output direction.

Although the distance functions defined by (2.4)–(2.5) are defined in terms of the production set $\mathcal{P}$, different distance functions can be defined by replacing $\mathcal{P}$ with some other set to measure distance from $(\boldsymbol{x}, \boldsymbol{y})$ to the boundary of the other set. For example, let $\mathcal{V}(\mathcal{A})$ denote the convex cone (with vertex at the origin) spanned by the set $\mathcal{A} \subset \mathbb{R}_+^{p+q}$. Clearly, $\mathcal{P} \subseteq \mathcal{V}(\mathcal{P})$. If $\mathcal{P}^\partial$ exhibits constant returns to scale (CRS) everywhere, then the technology $\mathcal{P}^\partial$ implies a mapping $\boldsymbol{x} \to \boldsymbol{y}$ that is homogeneous of degree 1; *i.e.,* $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}^\partial$ implies $(\lambda\boldsymbol{x}, \lambda\boldsymbol{y}) \in \mathcal{P}^\partial$ for all $\lambda > 0$. In this case, $\mathcal{P} = \mathcal{V}(\mathcal{P})$ and $\mathcal{I}(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}) = \mathcal{I}(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{V}(\mathcal{P})) = \mathcal{O}(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P})^{-1} = \mathcal{O}(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{V}(\mathcal{P}))^{-1}$ by construction. Otherwise, $\mathcal{P} \subset \mathcal{V}(\mathcal{P})$, $\mathcal{I}(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}) \leq \mathcal{I}(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{V}(\mathcal{P}))$, and $\mathcal{O}(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}) \geq \mathcal{O}(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{V}(\mathcal{P}))$.[1]

[1]Although not done here, it is possible to test the null hypothesis of constant returns to scale against the alternative hypothesis of variable returns to scale. See Simar and Wilson (2001) for details.

Of course, the production set $\mathcal{P}$ and hence the distance functions defined by (2.4)–(2.5) are unobserved and must be estimated from data. Before anything can be estimated, a statistical model must be defined. To ensure consistent estimation using the DEA estimator described below, Assumptions 2.1–2.3 listed above and the following assumptions of Kneip et al. (1998) are required:

Assumption 2.4. *The sample observations, $\mathcal{S}_n = \{(\boldsymbol{x}_i, \boldsymbol{y}_i\}_{i=1}^n\}$, are realizations of identically, independently distributed (iid) random variables with probability density function $f(\boldsymbol{x}, \boldsymbol{y})$ with support over $\mathcal{P}$.*

Assumption 2.5. *The density $f(\boldsymbol{x}, \boldsymbol{y})$ is continuous except along the frontier, with $f(\boldsymbol{x}, \boldsymbol{y}) = 0 \;\forall\; (\boldsymbol{x}, \boldsymbol{y}) \notin \mathcal{P}$ and $f(\boldsymbol{x}, \boldsymbol{y}) > 0 \;\forall\; (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}^{\partial}$.*

Assumption 2.6. *For all $(\boldsymbol{x}, \boldsymbol{y})$ in the interior of $\mathcal{P}$, $\mathcal{I}(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P})$ and $\mathcal{O}(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P})$ are differentiable in both their arguments.*

If panel data are observed, Assumption 2.4 may not hold. This does not affect consistency of non-parametric frontier estimators, but may change how one would use bootstrap methods for inference-making purposes; see Simar and Wilson (1999) for an example using Malmquist indices. Assumption 2.5 means that there is not a probability mass along the frontier, but the density of inputs and outputs is strictly positive along the frontier to ensure some non-zero probability of observing points in the sample near the frontier. Assumption 2.6 is stronger, but simpler, than the one used by Kneip et al. (1998); both are assumptions about the smoothness of the frontier $\mathcal{P}^{\partial}$. Together, Assumptions 2.1–2.6 define a statistical model. Estimation of quantities of interest is discussed in the next section.

3 Estimation and Inference

Several estimators of $\mathcal{P}$, and hence the distance functions $\mathcal{I}(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P})$ and $\mathcal{O}(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P})$, are possible. Deprins et al. (1984) proposed the FDH of the observations in $\mathcal{S}_n$, *i.e.*,

$$\widehat{\mathcal{P}}_{\text{FDH}}(\mathcal{S}_n) = \bigcup_{(\boldsymbol{x}_i, \boldsymbol{y}_i) \in \mathcal{S}_n} \{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_+^{p+q} \mid \boldsymbol{y} \le \boldsymbol{y}_i,\; \boldsymbol{x} \ge \boldsymbol{x}_i\}. \tag{3.1}$$

DEA estimators are obtained by replacing $\mathcal{P}$ in (2.4)–(2.5) with the convex hull of $\widehat{\mathcal{P}}_{\text{FDH}}$, given by

$$\widehat{\mathcal{P}}_{\text{VRS}}(\mathcal{S}_n) = \Big\{(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{R}_+^{p+q} \mid \boldsymbol{y} \leq \sum_{i=1}^{n} \gamma_i \boldsymbol{y}_i,\ \boldsymbol{x} \geq \sum_{i=1}^{n} \gamma_i \boldsymbol{x}_i, \sum_{i=1}^{n} \gamma_i = 1,\ \gamma_i \geq 0\ \forall\ i = 1,\ \ldots,\ n\Big\}. \tag{3.2}$$

Under the assumptions given in section 2, both $\widehat{\mathcal{P}}_{\text{FDH}}(\mathcal{S}_n)$ and $\widehat{\mathcal{P}}_{\text{VRS}}(\mathcal{S}_n)$ are consistent estimators of the production set $\mathcal{P}$. Asymptotic properties of $\widehat{\mathcal{P}}_{\text{FDH}}(\mathcal{S}_n)$ and $\widehat{\mathcal{P}}_{\text{VRS}}(\mathcal{S}_n)$ have been examined by Korostelev et al. (1995a, 1995b); see Simar and Wilson (2000b) for a summary.

Estimators of the distance functions $\mathcal{I}(\boldsymbol{x},\boldsymbol{y} \mid \mathcal{P})$ and $\mathcal{O}(\boldsymbol{x},\boldsymbol{y} \mid \mathcal{P})$ defined in (2.4)–(2.5) are obtained by replacing $\mathcal{P}$ with either $\widehat{\mathcal{P}}_{\text{FDH}}(\mathcal{S}_n)$ or $\widehat{\mathcal{P}}_{\text{VRS}}(\mathcal{S}_n)$ in (2.4)–(2.5). Variable returns-to-scale DEA estimators $\mathcal{I}(\boldsymbol{x},\boldsymbol{y} \mid \widehat{\mathcal{P}}_{\text{VRS}}(\mathcal{S}_n))$ and $\mathcal{O}(\boldsymbol{x},\boldsymbol{y} \mid \widehat{\mathcal{P}}_{\text{VRS}}(\mathcal{S}_n))$ are easily computed by linear programming methods, while the FDH estimators $\mathcal{I}(\boldsymbol{x},\boldsymbol{y} \mid \widehat{\mathcal{P}}_{\text{FDH}}(\mathcal{S}_n))$ and $\mathcal{O}(\boldsymbol{x},\boldsymbol{y} \mid \widehat{\mathcal{P}}_{\text{FDH}}(\mathcal{S}_n))$ can be computed by simple numerical methods. The estimators based on the convex hull permit varying returns to scale, and incorporate the convexity assumption in Assumption 2.1. The FDH estimators do not incorporate convexity, and remain consistent if the production set $\mathcal{P}$ is not convex.

Asymptotic properties of the DEA and FDH distance function estimators are discussed in Gijbels et al. (1999), Park et al. (2000), Simar and Wilson (2000b), and Kneip et al. (2003). In particular, it is now known that

$$\mathcal{I}(\boldsymbol{x},\boldsymbol{y} \mid \widehat{\mathcal{P}}_{\text{VRS}}) = \mathcal{I}(\boldsymbol{x},\boldsymbol{y} \mid \mathcal{P}) + O_p\left(n^{-2/(p+q+1)}\right) \tag{3.3}$$

and

$$\mathcal{I}(\boldsymbol{x},\boldsymbol{y} \mid \widehat{\mathcal{P}}_{\text{FDH}}) = \mathcal{I}(\boldsymbol{x},\boldsymbol{y} \mid \mathcal{P}) + O_p\left(n^{-1/(p+q)}\right), \tag{3.4}$$

and similarly for the corresponding output-oriented estimators. The convergence rates are slow, reflecting the curse of dimensionality which is common with non-parametric estimators.[2] The rate of convergence for the FDH estimator is slower than for the DEA estimator, but if $\mathcal{P}$ is non-convex, the

[2]The curse of dimensionality results here from the fact that the convergence rates in (3.3) and (3.4) depend on the sum of numbers of inputs and outputs $(p+q)$.

DEA estimator is statistically inconsistent. In addition to slow convergence rates and the curse of dimensionality, the DEA and FDH estimators also suffer from extreme sensitivity to outliers; *e.g.,* see Wilson (1995). For many applications, these problems are potentially acute.

For purposes of estimating scale efficiency, the convex cone of $\widehat{\mathcal{P}}_{\mathrm{VRS}}(\mathcal{S}_n)$,

$$\widehat{\mathcal{P}}_{\mathrm{CRS}}(\mathcal{S}_n) = \Big\{(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{R}_{+}^{p+q} \mid \boldsymbol{y} \leq \sum_{i=1}^{n} \gamma_i \boldsymbol{y}_i,\ \boldsymbol{x} \geq \sum_{i=1}^{n} \gamma_i \boldsymbol{x}_i,\ \gamma_i \geq 0\ \forall\ i = 1,\ \ldots,\ n\Big\}, \tag{3.5}$$

incorporates constant returns-to-scale. In addition,

$$\widehat{\mathcal{P}}_{\mathrm{NIRS}}(\mathcal{S}_n) = \Big\{(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{R}_{+}^{p+q} \mid \boldsymbol{y} \leq \sum_{i=1}^{n} \gamma_i \boldsymbol{y}_i,\ \boldsymbol{x} \geq \sum_{i=1}^{n} \gamma_i \boldsymbol{x}_i,\ \sum_{i=1}^{n} \gamma_i \leq 1,\ \gamma_i \geq 0\ \forall\ i = 1,\ \ldots,\ n\Big\}, \tag{3.6}$$

incorporates non-increasing returns to scale. By construction, $\widehat{\mathcal{P}}_{\mathrm{VRS}}(\mathcal{S}_n) \subseteq \widehat{\mathcal{P}}_{\mathrm{NIRS}}(\mathcal{S}_n) \subseteq \widehat{\mathcal{P}}_{\mathrm{CRS}}(\mathcal{S}_n)$. By comparing distance function estimates based on (3.2), (3.5), and (3.6), it is possible to determine whether the projection of an observation onto the estimated variable-returns frontier lies in a region of increasing, constant, or decreasing returns to scale.

Although Kneip et al. (2003) have derived sampling distributions for the DEA distance function estimators, these distributions are complicated and involve unknown parameters that must be estimated. As a practical matter, bootstrap methods remain the only useful approach to inference. These methods are summarized by Simar and Wilson (2000b). In this study, the homogeneous bootstrap introduced by Simar (1998) is used, and confidence intervals are estimated from bootstrap values using the percentile method described in Simar and Wilson (2000a), which incorporates an implicit correction for the bias in DEA distance function estimators.

4 The Data

4.1 The 2000 PISA database

Data for this project were taken from the student and school files contained in the 2000 PISA database. Details on PISA can be found on the internet at http://www.pisa.oecd.org/.

4.2 Specification of inputs and outputs

After considering what might be reasonably measured using the PISA data, three broad input categories were identified: socio-economic status, aggregate physical inputs, and aggregate teacher quality. The following list shows variables from the 2000 PISA database that might be used to measure each of the three inputs:

1. Socio-economic status:

   hdres Index of home educational resources, derived from students' reports on the availability in their home of a dictionary, a quiet place to study, a desk for study, textbooks, and the number of calculators in the home.

   wealth Index of family wealth, derived from students' reports on the availability in their home of a dishwasher, a room of their own, educational software, a link to the internet, and the number of cellular phones, television sets, computers, motor cars and bathrooms at home.

2. Aggregate physical inputs:

   teach number of full-time teachers plus one-half times number of part-time teachers.

3. Aggregate teacher quality:

   propcert proportion of certified teachers.

Variables **hdres** and **wealth** are observed directly in the student file, while **propcert** is observed directly in the school file. The aggregate physical input variable **teach** was computed from variables in the school file. In terms of the variables in the school file, **teach** was computed by dividing number of students (schlsize) by the student-teacher ratio (stratio).

Three output categories were also identified: test scores, grade level attainment, and number of students. The following list shows variables that might be used to measure each of these output categories:

1. Test scores:

   wlemath Warm estimate in mathematics.

   wleread Warm estimate in reading.

   wleread1 Warm estimate in reading—retrieving.

   wleread2 Warm estimate in reading—interpreting.

   wleread3 Warm estimate in reading—reflecting.

   wlescie Warm estimate in science. problem-solving.

2. Grade level attainment

   grade grade level attained at age 15 years.

3. Number of students

   students Number of students in each school.

Warm estimates in mathematics, reading, and science reflect statistical adjustments to test scores to allow comparison across countries. These are described in technical documentation available from PISA at the internet address given above.

The PISA database contains school-level observations as well as observations on individual students. In particular, the variables measuring socioeconomic status, test scores, and grade-level attainment are observed for each student in the survey, while variables describing number of teachers, proportion of certified teachers, and number of students are observed at the school level. For purposes of this study, individual schools are the unit of observation. Consequently, data at the student level were averaged by schools to obtain school-level observations.

After deleting observations for which at least one of the variables listed above could not be computed due to missing values, and further deleting observations with obviously incorrect values, 5,528 observations remain, representing schools in 40 of the 43 countries participating in PISA in 2000.

Table 1 lists these 40 countries, with the number of schools from each country that are represented in the final sample of 5,528 observations.[3] The 40 countries included in the final sample include eight former Soviet states (Albania, Bulgaria, Czech Republic, Latvia, Macedonia, Poland, Romania, and Russia), five Latin American countries (Argentina, Brazil, Chile, Mexico, and Peru), and four less-developed East Asian countries (Hong Kong, Indonesia, South Korea, and Thailand).

In order to reduce the number of input-output specifications and to avoid excessive numbers of input and output quantities that would exacerbate the slow convergence rate of DEA estimators, the sample (Pearson) correlation coefficient for the two socio-economic variables was computed as 0.7863. Eigenvalues λ_j, $j = 1, \ldots, 2$ of the (2×2) correlation matrix were computed and then normalized by computing $\lambda_j^* = \lambda_j / \sum_{k=1}^{2} \lambda_k$. Then $\sum_{j=1}^{2} \lambda_j^* = 1$, and each $\lambda_j^* \in [0, 1]$ gives the proportion of the independent linear information contained in each of the five corresponding principal components. In the present case, ordering the normalized eigenvalues (and the corresponding eigenvectors) so that $\lambda_1^* \geq \lambda_2^*$ reveals that the first principal component of the two socio-economic variables contains 89.32-percent of the independent linear information in these variables.[4] Consequently, a new variable, **social**, was defined as the first principal component of **hdres** and **wealth** to be used in the efficiency analysis of PISA schools.

The variables **wlemath**, **wleread**, **wleread1**, **wleread2**, **wleread3**, and **wlescie** describing test scores in mathematics, reading, and science are also highly correlated with each other; their Pearson correlation matrix appears in Table 2. The first principal component accounts for 89.57-percent of the independent linear information contained in these variables, with elements of the corresponding eigenvector ranging from 0.374 to 0.429; consequently, each variable contributes roughly equally to the first principal component, as one might expect from the correlations in Table 2 which are all close to unity. A new variable, **test**, was defined as the first principal component of the six variables **wlemath**, **wleread**, **wleread1**, **wleread2**, **wleread3**, and

[3]Three-character country names in Table 1 conform to ISO 3166 standards.

[4]Given an $(N \times K)$ matrix $\boldsymbol{X}$ containing N observations on K variables, it is straightforward to compute eigenvectors given by the K columns of the $(K \times K)$ matrix $\boldsymbol{E}$. Of course, the eigenvalues and their corresponding eigenvectors can be ordered so that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_5$, with the eigenvector in the first column of $\boldsymbol{E}$ corresponding to λ_1, *etc.* Then the first principal component of the data in $\boldsymbol{X}$ is given by the $(N \times 1)$ matrix $\boldsymbol{X}\boldsymbol{E}_{\cdot 1}$, where $\boldsymbol{E}_{\cdot 1}$ denotes the first column of $\boldsymbol{E}$.

wlescie; this variable will be used to measure academic performance in the efficiency analysis that appears later.

4.3 Data characteristics

Table 3 gives summary statistics for the final input and output variables described above. In several cases, means and medians differ, suggesting skewed distributions. This is confirmed in Figure 1, where non-parametric kernel density estimates for each input and output are plotted.[5] In particular, the distributions for **teach** and **students** have long right tails; there are many very small schools in the sample, as well as a relatively small number of much larger schools. Similar phenomena are observed for **propcert**, whose distribution has a long left tail. In other words, the vast majority of teachers are certified.

It is well-known that envelopment estimators such as DEA and FDH are sensitive to outliers. The data were analyzed using the influence-function approach of Wilson (1993), but the results were inconclusive. This appears to be due to the fact that the data for three variables are heavily skewed, as discussed above. Consequently, a very large number of observations appear as outliers when the Wilson (1993) diagnostic is used. The existence of outliers, however, does not necessarily mean that data have somehow been corrupted (*e.g.,* through coding errors, mis-reporting, *etc.*). In the present case, given the heavily skewed distributions for three of the six variables of interest, one should expect to find numerous outliers. The summary statistics in Table 3 indicate, however, that the values for each observation are not implausible.

To put things differently, the fact that there are outliers in the data appears to result from distributional features of the data. In a certain sense, the outliers may be some of the most important observations in the data. When drawing observations from a heavily skewed distribution, for example, one with a long right tail, one will typically draw small values of the deviate and will only infrequently draw large values. Consequently, one would have to draw many times to learn much about the features of the distribution, in particular information about the long right tail. The plots of estimated

[5]The kernel density estimates shown in Figure 1 were obtained using a Gaussian kernel function and bandwidths selected using the two-stage plug-in procedure of Sheather and Jones (1991). In cases where the support of a given distribution is bounded (*i.e.,* in all cases except **test**), the reflection method described by Silverman (1986) and Scott (1992) was used to avoid bias problems near the boundaries.

densities in Figure 1 suggest that the underlying marginal distributions are indeed heavily skewed. This, together with the fact that the sample size is reasonably "large"—5,528 observations—apparently means that the sample size is large enough to reveal information about distributional features that are typically difficult to uncover.

5 Results of Efficiency Estimation

5.1 Average (and median) efficiency by country

Initially, DEA as well as FDH estimates of the input and output distance functions $\mathcal{I}(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P})$ and $\mathcal{O}(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P})$ defined in (2.4)–(2.5) were computed for each of the 5,528 schools in the sample using the three inputs (**social**, **teach**, and **propcert**) and the three outputs (**test**, **grade**, and **students**) defined in section 4.2.

Table 4 shows mean, as well as median, input-oriented DEA and FDH efficiency estimates for each of the 40 countries represented in the sample, as well as mean and median estimates for the entire sample in the last row of the table. Country abbreviations are given in the first column; the results have been ordered by mean input-efficiency estimates, from most-efficient to least-efficient. Columns 2–3 give the mean input-oriented DEA and FDH efficiency estimates, and columns 5–6 give the median input-oriented estimates for each country. The last column gives the number of observations in the sample for each country.

Comparison of the FDH and DEA estimates serves as a useful diagnostic. Recall from the discussion in Section 2 that the DEA estimator of the production set is the convex hull of the FDH estimator of the production set. The results in Table 4 indicate that many of the FDH efficiency estimates are different from unity. If this were not the case—*i.e.,* if most or all of the FDH estimates were equal to one—this would indicate that any apparent inefficiency indicated by the DEA estimator would merely be a consequence of the convexity assumption maintained by the DEA estimator. This problem can arise when the curse of dimensionality that was described in Section 2 is critical. Fortunately, the results in Table 4 indicate that this is not the case in the present application.

Columns 4 and 7 in Table 4 give the mean and median number of observations that weakly dominate each school in each country. Observation j

weakly dominates observation i if $x_j \leq x_i$ and $y_j \geq y_i$, or in other words if observation j uses weakly less of all inputs and yet produces weakly greater amounts of each output than observation i. For every case in Table 4, the mean number of dominating schools exceeds the median number of dominating schools, suggesting that the *distribution* of dominance relationships is skewed, with a long right-tail. Many schools are weakly dominated either only by themselves (by definition, each observation weakly dominates itself), or by perhaps a small number of other schools, while a smaller number of schools are weakly dominated by many other schools.

The dominance results in Table 4 give an idea of the "depth" of the data. In the case of Romania (ROM), which appears as the second-least efficient country in terms of average input efficiency of its schools, each school is weakly dominated by, on average, 172 schools. This means that on average, 171 dominating peers can be found for each Romanian school in the sample. An inefficient school in Romania may have little to learn from relatively more efficient schools in the sample that are very different in terms of their compositions of inputs and outputs, but the dominating peers for a given school provide poignant comparisons: the dominating peers produce *at least* as much of each output while using *no more* than the observed level of each input for the given school.

In terms of overall, country-wide performance, among the transitioning countries represented in the sample, average input efficiency ranges from 2.0370 for Macedonia (MKD) to 2.9345 for Romania. The transitioning economies seem to fall into two groups; average input-efficiencies are similar for Russia (RUS), Latvia (LVA), Bulgaria (BLG), Poland (POL), and Romania (ROM), while average input-efficiencies are also similar, and higher than for the first group, for Macedonia (MKD), the Czech Republic (CZE), and Albania (ALB).

Similar to Table 4, Table 5 gives mean and median output-oriented efficiency estimates by country. As in Table 4, results have been sorted by the mean output efficiencies, from most-efficient to least-efficient. The average of 0.8074 for Romania in Table 5 suggests that outputs in Romanian schools are on average only 0.8074 times thier (estimated) efficient levels. In other words, if technical efficiency were completely eliminated in Romanian schools, the result suggests that outputs could be increased by a factor of $1/0.8074 \approx 1.2385$—about 24-percent—without increasing input quantities. Again, this issue is examined further in Section 5.6 below.

The results in Table 5 reveal that in the output orientation, mean and median efficiency estimates are closer to one than was the case in the input orientation. For a number of countries, median FDH efficiency estimates are equal to one (this also occurs in Table 4, but with much less frequency). This phenomenon is apparently due to inclusion of the output variable **grade**, representing grade-level attainment at age 15. Recall from Table 3 that this variable has a median value of 10.0, and a maximum observed value of 11.0. Hence large numbers of observations are at or near the maximum in terms of this variable. Consequently, in the output orientation, many schools appear highly efficient, but this is due to the constraint imposed by this single variable. In terms of the linear programs used to compute the distance function estimates, in many cases the constraint for **grade** is binding at the solution while constraints for the other outputs have slack.[6]

To further examine this problem, efficiency estimates were re-computed for both the input as well as the output-orientation while deleting **grade** from the analysis, leaving three inputs ($p = 3$) and two outputs ($q = 2$). Deleting **grade** produced little change in the input-oriented estimates, but rather larger changes in the output-orientation. Results for the output orientation, omitting **grade**, are shown in Table 6, where again results have been sorted by mean efficiencies.

Comparing the results in Tables 5 and 6 reveals that omitting **grade** reduces average (as well as median) output efficiency for each country. Among the transitioning countries, average efficiency for the Czech Republic is reduced least when **grade** is omitted, while Albania suffers the worst decline.

Which set of results for the output orientation are most meaningful? By construction, adding an output variable where all observations are equal to a constant would result in all output efficiency estimates equal to unity. However, this is not the case with **grade**, which has some variation as indicated by its sample standard deviation reported in Table 3. The kernel density estimate for **grade**, shown in Figure 1, reveals large modes at 9 and 10, with smaller modes at 8 and 11, and an even smaller mode at 7. When **grade** is included in the list of outputs, The results in Tables 4 and 5 suggest that

[6] A number of published studies in the economics and management science literatures have devoted excessive amounts of ink to discussions of slack in linear programs used to compute DEA efficiency estimates. From a statistical viewpoint, slack is merely a small-sample phenomenon and has little, if any interest. The discussions in the literature seem to ignore all statistical aspects of the problem, and imply that *truth* can be learned from *data.* This is utter nonsense, as can be seen by considering what might happen if a new observation is obtained.

most observations in the sample are relatively more efficient in the output-orientation than in the input-orientation. In short, the finding that observations are typically more output-efficient than they are input-efficient does not, by itself, mean that grade-level attainment (**grade**) should be deleted from the analysis. Consequently, the analysis that follows includes **grade**.

5.2 Performance of "typical" schools

The median and average efficiency results shown in Tables 4, 5, and 6 are obtained from point estimates. Of course, these are different from *true* efficiencies; they are only *estimates.* While the bootstrap methods mentioned near the end of Section 3 can be used to estimate confidence intervals for individual schools, it is informative to first consider the performance of "typical" schools in each country.

To give a concise idea of how reliable individual point estimates might be, 40 hypothetical observations were formed by taking the medians of each input and output quantity by country. These hypothetical, median schools may be viewed as the "typical" school in each country; rather than examining average and median performance of schools in each country as in section 5.1, one can examine the performance of the "typical" school in each country.

Efficiency of these hypothetical, median observations was estimated, relative to the frontier estimate based on the 5,528 original sample observations. Bootstrap methods were then used to estimate the bias variance and variance of these estimates, and to estimate confidence intervals for the efficiencies of the 40 hypothetical observations.

Results for the input-oriented version of this exercise are given in Table 7; these results were obtained with grade-level attainment (**grade**) included as an output. The first two columns of Table 7 give the country identifier and the (DEA) estimate of input efficiency for each country's hypothetical, median school. The next two columns give the bootstrap bias and standard deviation estimates, which are reasonably small as one might expect from 5,528 observations and $p + q = 6$ dimensions. The fifth column gives a bias-corrected estimate of efficiency, obtained by subtracting the bootstrap bias estimate from the original efficiency estimate. By construction, bias is negative when the input orientation is used, and so the bias-corrected estimates are larger (farther from unity) than the original estimates, indicating greater inefficiency than the original estimates.

The last two columns of Table 7 give lower and upper bounds for 95-

percent confidence interval estimates for efficiency of median schools in each country. The countries in Table 7 have been sorted by the lower bounds of their corresponding confidence interval estimates. The interval estimates are narrow, which might also be expected given the large sample size, but nonetheless some intervals overlap. Consequently, the rank-ordering by lower bounds of confidence interval estimates provides only a partial ordering. Given the narrowness of the confidence interval estimates, the results suggest that efficiency is estimated rather well.[7]

The results in Table 7 also give additional insight into performance across the nine countries of interest in this study. Comparing Tables 4 and 7, one sees some differences in the rankings. The estimated confidence intervals in Table 7 suggest that among the "typical" schools in the transitioning countries, Albania (ALB) and Macedonia (MKD) are similar, but better than the Czech Republic (CZE), which in turn is better than Russia (RUS) and Romania (ROM), whose confidence intervals overlap. Estimated confidence intervals for the three lowest-ranked transitioning countries in Table 7—Bulgaria (BGR), Poland (POL), and Latvia (LVA)—are distinct and do not overlap.

Similar to Table 7, Table 8 gives results of output-efficiency estimated for the 40 "typical" schools representing each country.[8] In Table 8, countries have been sorted by the upper bounds on their corresponding confidence interval estimates. Here, the estimated confidence intervals in the last two columns lie *below* the original point estimates in the second column of the table, due to the upward (positive) bias of the DEA output efficiency estimator. Although the rankings in Table 8 differ from those in Table 7, among the transitioning economies, the same four countries—Bulgaria, Poland, Latvia, and Romania—rank lowest in terms of their "typical" schools.

5.3 Individual school performance

As noted above in Section 5.2, the same bootstrap methods used to estimate confidence intervals for the hypothetical, median schools in each country can be used to estimate confidence intervals for individual schools in the sample.

[7]Note that the lower bounds of the confidence interval estimates in Table 7 lie *above* the original distance function estimates. This is due to the fact that the original estimates are biased downward; the bootstrap method used to estimate the confidence intervals involves an implicit correction for bias. See Simar and Wilson (2000b) for details.

[8]Entries such as "2.663e-05" in the second row of Table 8 use scientific notation; i.e., 2.663e-05 denotes 2.663×10^{-5}.

Rather than overwhelming the reader with many pages of numbers, however, the results in this section are presented graphically.

The bootstrap methods described near the end of Section 3 were used to estimate 95-percent confidence intervals for both input and output efficiency for each school in the eight transitioning countries included in the sample. Figures 2 and 3 show the results for input efficiency of schools in each of the transitioning countries. Confidence intervals for each school are represented by vertical line segments whose endpoints, measured on the vertical axis in each graph, correspond to upper and lower bounds of estimated confidence intervals. Within each country, observations have been sorted by the lower bounds of the confidence interval estimates. Thus, the horizontal axis in each graph in Figures 2 and 3 gives the rank of each school (from 1 to n_k, where n_k is the number of observations for country k) in terms of the lower bounds of the estimated confidence intervals.

The results in Figures 2 and 2 are of course consistent with the results on average and median input efficiency discussed in Section 5.1. Figures 2 and 4, however, give an idea of the *distribution* of inefficiency within each country. For example, schools in Romania, which was found to be among the worst performers in terms of average or median input efficiency in Table 4, do not universally perform worse than schools in other countries. The best-performing schools in Romania appear to perform better than perhaps half the schools in Albania, but the graph for Romania in Figure 3 indicates that performance falls off quickly for Romanian schools once one moves beyond the top-ranked schools. Performance of Polish schools also diminishes rapidly, but only for schools ranked beyond about 40–50. Schools in the Czech Republic and in Macedonia appear rather homogeneous in terms of their performances.

Figures 4 and 5 display confidence interval estimates for output efficiency (with the grade-level attainment variable, **grade**, included). In these figures, observations for each country have been sorted by the upper bound of the confidence interval estimates, so that the horizontal axis again gives the rank of each school in terms of the intervals' upper bounds.

It is interesting to note that while Table 5 suggested that average and median efficiency estimates for schools in each country were rather high, many of the confidence intervals shown in Figures 4 and 5 are rather wide, particularly for schools in Albania, Macedonia, and Poland. In addition, there appears to be some grouping among the observations within each country. Latvia provides the most pronounced example of this; the groups are due to the grouping in the output variable **grade** that was discussed earlier (*e.g.,* see the kernel density estimate for **grade** in Figure 1 showing multiple modes).

5.4 Returns to scale

As discussed in Section 2, constant-returns to scale and non-increasing-returns to scale versions of both the input and output DEA estimators can be used to determine whether observations projected onto the frontier (in either the input or output directions) would face increasing, constant, or decreasing returns to scale. It is important to note that returns to scale is a property of the frontier. Hence it makes no sense to speak of returns to scale for points that are not on the frontier, but in the interior of the set of feasible input/output combinations.

Table 9 gives the results of such an exercise. The results indicate that when observations are projected onto the frontier using the input-efficiency estimates to scale back inputs to technically efficient levels, more than two-thirds of Bulgarian, Macedonian, and Romanian ier, while a majority of schools in Argentina and almost all schools in Hong Kong and South Korea would lie on the part of the frontier where returns to scale are decreasing.

In the output orientation, however, Table 9 confirms that if outputs were scaled up to efficient levels using the DEA output efficiency estimates, almost all schools in the sample would lie on the decreasing-returns portion of the frontier, as suggested by the discussion of output efficiency in Section 5.1. In terms of the one input, one output example depicted in Figure

5.5 Differences in input/output mixes

In order to examine whether relatively more-efficient schools use different mixes of inputs or outputs than relatively less-efficient schools, schools in the transitioning countries represented in the sample were sorted first by their DEA estimates of input efficiency and then by their DEA estimates of output efficiency. In the input orientation, the first quartile lies at 2.0259, while in the output orientation, the third quartile lies at 0.9090.

Observations from the eight transitioning countries were divided into two groups according to whether their DEA estimates of input efficiency were greater than 2.0259 or less than or equal to 2.0259. Mean values of input and output quantities used by schools in these two groups are reported in

Table 10; 837 schools fall in the less-efficient group, while there are 279 schools in the more-efficient group.

Table 10 reveals that means for each of the three input variables are greater for the less-efficient schools than the corresponding means for the more-efficient schools. describing socio-economic status (**social**) and number of teachers (**teach**) are almost identical across the two groups. In addition, means for each of the three output variables are smaller for the less-efficient schools than for the more-efficient schools. Standard central-limit theorem reasoning suggests that the differences are significant in each case at better than .01 significance.[9]

Next, schools in the transitioning countries were divided into two groups according to whether their output efficiency estimates were less than 0.9090 or weakly greater than 0.9090. Mean values of input and output quantities used by schools in these two groups are reported in Table 11; 669 schools fall in the less-efficient group, while there are 447 schools in the more-efficient group.[10] Table 11 gives mean input and output levels among schools in the two groups.

Table 11 reveals phenomena similar to those found in Table 10 for the input-orientation. In Table 11, as in Table 10, mean levels each input variable are higher for the less-efficient schools than for the more-efficient schools, while the means of the output variables are higher for the more-efficient schools than for the less-efficient schools. As in the input-oriented case, central limit theorem reasoning suggests that differences in corresponding means across the two groups are statistically significant at better than .01 significance.

5.6 Potential gains from elimination of inefficiency

As discussed in Section 2, the input distance function gives the maximum, feasible, proportionate reduction in input levels for a given school. Similarly, the output distance function gives the maximum, feasible, proportionate increase in output levels for a given school. These can be estimated by dividing input and output quantities observed for each school by the corresponding estimates of input and output distance functions.

[9] These results, however, are at best suggetive, due to the problems discussed in Simar and Wilson (2005).

[10] Although 0.9090 is the third-quartile of the output efficiency estimates, many estimates are equal to this value due to the clustering in **grade** that was discussed earlier.

Columns 2–4 in Table 12 give, for each country in the sample, estimates of the potential percentage reductions in input quantities that would result from elimination of all technical inefficiency. Columns 5–7 give estimates of the potential percentage increases in output quantities that would result from eliminating technical efficiency. These values were determined by computing, for each school in a given country, the ostesibly efficient levels of inputs and outputs by dividing observed input quantities for a given school by that school's estimated input distance function, and its observed output quantities by its estimated output distance function (recall the discussion on page 3). Summing across schools yields total observed input and output quantities, as well as total (estimated) efficient levels of inputs and outputs. The percentatges in Table 12 reflect the difference between total observed input (or output) and total efficient input (or output), divided by total observed input (or output) (the resulting quanitities are then multiplied by 100 to obtain percentages). The estimates in the two panels of Table 12 assume that technical inefficiency is eliminated either by scaling input quantities back to their efficient levels, or by scaling output quantities up to their efficient levels, but not by both.

Consistent with the high levels of average and median input inefficiency discussed in Section 5.1, rather large levels of savings are possible in the input direction, with rather smaller potential increases in the output direction. For example, the results suggest that Romania could, in principle, reduce its number of teachers by as much as 64.44 percent from the observed level without reducing output quantities. This saving of resources, however, would only be possible if educational services in Romania were managed efficiently.

The potential savings and gains suggested by the results in Table 12 reflect the radial nature of the input and output distance functions that were estimated. From the viewpoint of educational authorities in each country, potential reductions in numbers of teachers or proportions of certified teachers may be of interest, but perhaps not social status (measured by the variable **social**). Students' social status is, in most (perhaps all) cases, outside the control of school managers.

6 Final Remarks

This paper has analyzed efficiency in provision of education services among countries represented in the year 2000 PISA survey, with emphasis on transitioning countries in the former Soviet Union. In broad terms, input inefficiency was found to be rather high relative to output inefficiency, suggesting that many schools operate under regions of the production frontier where returns to scale are decreasing.

Some caution should be exercised in interpreting the results in this study. Any empirical study can only be as good as the data that are used. Although the PISA data are unusual in terms of the number of countries that one can compare, the survey design is perhaps not the best for purposes of this study. As noted in Section 1, PISA attempts to provide representative data on 15-year old students in each country. Since data have been aggregated across students to obtain school-level data, the relevant population of schools consists of schools with at least one 15-year old student. The schools in the data used for estimation in this project may be heterogeneous; this is a question for future analysis.

Another problem concerns the use of average test scores as an output. Conceivably, two schools might have similar *average* test scores, but might have very different *distributions* of test scores among their students. Concepts such as stochastic dominance can be used to compare differences in distributions and to perhaps rank distributions, but it is far from clear how these ideas might be incorporated into a model of productive efficiency. Additional, basic research is needed in this area.

Data problems are always present. The statistical methods used in this study are sound and well-understood, however. Given the importance of the questions raised in Section 1, one can do nothing else but use the best data available. The fact that those data might not be perfect does not mean that a study using those data cannot impart useful information. As with any empirical study, the careful reader will remember that one never learns *truth* from data—data can only provide *estimates* of truth. With finite samples, these are never the same thing.

Despite any data problems, the results of this study provide useful diagnostics on the performance of educational service providers in various countries. Particularly inefficient schools are candidates for further scrutiny to determine the reasons why these schools perform poorly. Section 5.1 suggested a method for finding relevant, plausible peers that could serve as role models for an inefficient school.

In some cases, inefficiency might result from managerial ineptitude, while in other cases apparent inefficiency might result from constraints outside authorities' direct control. Conceivably, one might regress the estimated efficiencies obtained in this study on variables describing economic and other conditions in each country or perhaps in regions within a country in an attempt to learn more about the causes of inefficiency. A large number of efficiency studies in the economics and management science literatures have done this, but a warning is appropriate: most, perhaps all of these studies are seriously flawed and make little sense. Simar and Wilson (2005) discuss the problem and provide some solutions, and should be consulted before attempting any second-stage regressions.

References

Deprins, D., Simar, L., and Tulkens, H. (1984), "Measuring labor inefficiency in post offices," In *The Performance of Public Enterprises: Concepts and Measurements,* ed. M. Marchand P. Pestieau and H. Tulkens, Amsterdam: North-Holland pp. 243–267.

Färe, Rolf (1988), *Fundamentals of Production Theory,* Berlin: Springer-Verlag.

Gijbels, I., E. Mammen, B.U. Park, and Simar, L. (1999), "On estimation of monotone and concave frontier functions," *Journal of the American Statistical Association,* 94, 220–228.

Kneip, A., Park, B., and Simar, L. (1998), "A note on the convergence of nonparametric DEA efficiency measures," *Econometric Theory,* 14, 783–793.

Kneip, A., Simar, L., and Wilson, P.W. (2003), "Asymptotics for DEA estimators in nonparametric frontier models," Discussion paper #0317, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Korostelev, A., Simar, L., and Tsybakov, A.B. (1995a), "Efficient estimation of monotone boundaries," *The Annals of Statistics,* 23, 476–489.

Korostelev, A., Simar, L., and Tsybakov, A.B. (1995b), "On estimation of monotone and convex boundaries," *Publ. Statist. Univ. Paris XXXIX,* 1, 3–18.

Park, B., Simar, L., and Weiner, C. (2000), "FDH efficiency scores from a stochastic point of view," *Econometric Theory,* 16, 855–877.

Scott, David (1992), *Multivariate Density Estimation: Theory, Practice, and Visualization,* New York: John Wiley & Sons, Inc.

Sheather, S.J., and Jones, M.C. (1991), "A reliable data-based bandwidth selection method for kernel density estimation," *Journal of the Royal Statistical Society, Series B,* 53, 683–690.

Shephard, Ronald W. (1970), *Theory of Cost and Production Functions,* Princeton: Princeton University Press.

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

Simar, L. and P.W. Wilson (1998), "Productivity growth in industrialized countries," Discussion Paper #9810, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Simar, L., and Wilson, P.W. (1999), "Estimating and bootstrapping malmquist indices," *European Journal of Operational Research,* 115, 459–471.

—— (2000a), "A general methodology for bootstrapping in non-parametric frontier models," *Journal of Applied Statistics,* 27, 779–802.

—— (2000b), "Statistical inference in nonparametric frontier models: The state of the art," *Journal of Productivity Analysis,* 13, 49–78.

—— (2001), "Nonparametric tests of returns to scale," *European Journal of Operational Research,* 139, 115–132.

—— (2005), "Estimation and inference in two-stage, semi-parametric models of productive efficiency," *Journal of Econometrics.* Forthcoming.

Wilson, Paul W. (1993), "Detecting outliers in deterministic nonparametric frontier models with multiple outputs," *Journal of Business and Economic Statistics,* 11, 319–323.

—— (1995), "Detecting influential observations in data envelopment analysis," *Journal of Productivity Analysis,* 6, 27–45.

Table 1: Countries and number of schools in final sample.

country	code	#	country	code	#
Albania	ALB	120	Israel	ISR	45
Argentina	ARG	118	Italy	ITA	134
Australia	AUS	200	Japan	JPN	125
Belgium	BEL	178	South Korea	KOR	143
Bulgaria	BGR	120	Liechtenstein	LIE	9
Brazil	BRA	195	Luxembourg	LUX	19
Switzerland	CHE	223	**Latvia**	LVA	107
Chile	CHL	125	Mexico	MEX	131
Czech Republic	CZE	210	Macedonia	MKD	60
Germany	DEU	187	Netherlands	NLD	69
Denmark	DNK	146	Norway	NOR	135
Spain	ESP	153	New Zeeland	NZL	142
Finland	FIN	138	Peru	PER	99
France	FRA	122	**Poland**	POL	113
Great Britain	GBR	287	Portugal	PRT	143
Greece	GRC	154	**Romania**	ROM	171
Hong Kong	HKG	132	**Russia**	RUS	215
Indonesia	IDN	232	Sweden	SWE	127
Ireland	IRL	136	Thailand	THA	165
Iceland	ISL	102	United States	USA	98

Table 2: Pearson correlation matrix for test-score variables.

	wlemath	**wleread**	**wleread1**	**wleread2**	**wleread3**	**wlescie**
wlemath	1.0000	0.8517	0.8571	0.8428	0.8123	0.6066
wleread	0.8517	1.0000	0.9828	0.9886	0.9695	0.8418
wleread1	0.8571	0.9828	1.0000	0.9607	0.9353	0.8327
wleread2	0.8428	0.9886	0.9607	1.0000	0.9432	0.8356
wleread3	0.8123	0.9695	0.9353	0.9432	1.0000	0.8131
wlescie	0.6066	0.8418	0.8327	0.8356	0.8131	1.0000

Table 3: Summary statistics for input and output variables.

Variable	mean	std. dev.	minimum	median	maximum
social	2.6390	1.1720	0.3242	2.2580	7.0770
teach	50.1400	36.9364	1.5000	42.0200	542.8000
propcert	0.8063	0.3087	0.0000	0.9700	1.0000
test	975.6000	161.3865	459.9000	997.4000	1390.0000
grade	9.5400	0.7366	7.0000	10.0000	11.0000
students	731.5000	592.5040	2.0000	606.0000	9815.0000

Table 4: Mean and median efficiency estimates by country, input orientation, $p = q = 3$.

	mean			median			
	DEA	FDH	dom	DEA	FDH	dom	n_k
GBR	1.3757	1.0679	9.7	1.2896	1.0101	2.0	287
NZL	1.4183	1.0959	3.2	1.4086	1.0173	2.0	142
MEX	1.5026	1.1758	4.2	1.4681	1.0000	2.0	131
CHL	1.5081	1.1489	5.0	1.5110	1.0766	3.0	125
ISL	1.5669	1.0613	3.0	1.6050	1.0000	2.0	102
BRA	1.6766	1.3135	62.0	1.4820	1.1618	5.0	195
DEU	1.7368	1.1012	11.8	1.6909	1.0326	3.0	187
NOR	1.7467	1.1525	7.9	1.7070	1.1023	4.0	135
NLD	1.8025	1.1706	24.1	1.6745	1.0526	2.0	69
AUS	1.8164	1.1817	28.0	1.8408	1.1155	12.0	200
HKG	1.9076	1.1934	20.6	1.8713	1.0863	6.0	132
KOR	1.9087	1.2084	32.7	1.8142	1.0417	5.0	143
JPN	1.9324	1.2267	31.7	1.9177	1.1692	12.0	125
USA	1.9400	1.2086	28.0	1.8845	1.1619	8.0	98
SWE	1.9422	1.1630	10.8	1.9365	1.1047	5.0	127
ISR	1.9524	1.2962	19.3	1.8988	1.2000	6.0	45
CHE	1.9692	1.2403	24.2	1.9359	1.1826	6.0	223
FIN	1.9703	1.1651	9.6	1.9966	1.1429	5.5	138
IRL	1.9987	1.2178	18.4	1.9772	1.1921	7.5	136
MKD	2.0370	1.4163	33.8	2.0056	1.4759	9.5	60
DNK	2.0449	1.2424	53.2	2.0743	1.1936	18.5	146
THA	2.0632	1.4691	46.1	1.9826	1.4074	12.0	165
CZE	2.0655	1.2883	9.4	2.0600	1.2685	5.0	210
PER	2.0713	1.5181	38.9	2.0089	1.4667	15.0	99
LIE	2.0808	1.2863	48.8	1.9842	1.2333	11.0	9
ALB	2.0820	1.3635	32.2	1.9396	1.0185	7.0	120
FRA	2.1036	1.2808	48.5	2.1031	1.2586	20.0	122
ESP	2.1128	1.2993	110.1	2.1573	1.3082	63.0	153
ITA	2.2445	1.3769	62.0	2.2565	1.3885	21.5	134
BEL	2.2965	1.4204	160.7	2.0773	1.2658	40.5	178
IDN	2.3708	1.6080	65.3	2.3012	1.5324	18.5	232
LUX	2.3759	1.2673	35.2	2.4458	1.2059	6.0	19
GRC	2.4631	1.5246	41.0	2.4159	1.4444	21.0	154
RUS	2.5669	1.6654	48.8	2.6003	1.6565	21.0	215
PRT	2.6973	1.5548	69.8	2.6814	1.5321	32.0	143
LVA	2.7927	1.8069	116.4	2.7501	1.7754	43.0	107
BGR	2.8886	1.8138	146.5	2.9032	1.7820	79.0	120
POL	2.9120	1.8932	270.5	2.7297	1.7715	68.0	113
ROM	2.9345	2.0405	172.0	2.8855	2.0425	77.0	171
ARG	3.2362	2.0952	318.9	3.2736	2.0773	199.0	118
TOTAL	2.0837	1.3648	56.1	1.9908	1.2249	9.0	5528

Table 5: Mean and median efficiency estimates by country, output orientation, $p = q = 3$.

	mean			median			
	DEA	FDH	dom	DEA	FDH	dom	n_k
NZL	1.0000	1.0000	3.2	1.0000	1.0000	2.0	142
GBR	0.9802	0.9920	9.7	1.0000	1.0000	2.0	287
ISL	0.9253	0.9881	3.0	0.9106	1.0000	2.0	102
AUS	0.9241	0.9540	28.0	0.9090	0.9458	12.0	200
NOR	0.9176	0.9788	7.9	0.9090	1.0000	4.0	135
JPN	0.9153	0.9595	31.7	0.9090	0.9740	12.0	125
KOR	0.9136	0.9815	32.7	0.9167	1.0000	5.0	143
HKG	0.9120	0.9781	20.6	0.9110	1.0000	6.0	132
ISR	0.9081	0.9571	19.3	0.9090	1.0000	6.0	45
ALB	0.9044	0.9740	32.2	0.9108	1.0000	7.0	120
ITA	0.9025	0.9206	62.0	0.9090	0.9091	21.5	134
ESP	0.9023	0.9327	110.1	0.9090	0.9091	63.0	153
GRC	0.9008	0.9474	41.0	0.9090	0.9614	21.0	154
MEX	0.8990	0.9667	4.2	0.9141	1.0000	2.0	131
RUS	0.8936	0.9528	48.8	0.9090	0.9917	21.0	215
BEL	0.8935	0.9145	160.7	0.9090	0.9091	40.5	178
CHL	0.8925	0.9620	5.0	0.9194	1.0000	3.0	125
USA	0.8889	0.9395	28.0	0.9090	0.9402	8.0	98
NLD	0.8858	0.9433	24.1	0.8968	0.9578	2.0	69
CZE	0.8821	0.9553	9.4	0.9090	0.9693	5.0	210
IRL	0.8765	0.9456	18.4	0.8842	0.9434	7.5	136
FRA	0.8700	0.9120	48.5	0.9090	0.9091	20.0	122
ARG	0.8694	0.8865	318.9	0.9090	0.9091	199.0	118
PER	0.8651	0.9308	38.9	0.8682	0.9091	15.0	99
THA	0.8602	0.9311	46.1	0.8351	0.9000	12.0	165
DEU	0.8532	0.9475	11.8	0.8475	0.9698	3.0	187
CHE	0.8520	0.9259	24.2	0.8314	0.9091	6.0	223
FIN	0.8502	0.9414	9.6	0.8485	0.9308	5.5	138
PRT	0.8470	0.8657	69.8	0.8307	0.9091	32.0	143
LVA	0.8448	0.8910	116.4	0.8183	0.9000	43.0	107
IDN	0.8433	0.9068	65.3	0.8225	0.9000	18.5	232
LIE	0.8416	0.9145	48.8	0.8430	0.9000	11.0	9
MKD	0.8395	0.9049	33.8	0.8285	0.9000	9.5	60
DNK	0.8387	0.8980	53.2	0.8254	0.9000	18.5	146
SWE	0.8386	0.9375	10.8	0.8291	0.9235	5.0	127
POL	0.8341	0.8774	270.5	0.8181	0.9000	68.0	113
LUX	0.8281	0.8783	35.2	0.8181	0.9000	6.0	19
BRA	0.8179	0.8895	62.0	0.8268	0.9000	5.0	195
BGR	0.8171	0.8593	146.5	0.8181	0.8476	79.0	120
ROM	0.8074	0.8559	172.0	0.8181	0.8536	77.0	171
TOTAL	0.8819	0.9350	56.1	0.9090	0.9291	9.0	5528

Table 6: Mean and median efficiency estimates by country, output orientation, $p = 3$, $q = 2$ (output **grade** deleted).

	mean			median			
	DEA	FDH	dom	DEA	FDH	dom	n_k
NLD	0.8298	0.9148	24.6	0.8622	0.9485	2.0	69
FIN	0.8202	0.9217	9.7	0.8193	0.9206	5.5	138
HKG	0.8138	0.9055	25.0	0.8278	0.9268	7.0	132
MEX	0.8099	0.9117	4.7	0.7999	0.9585	2.0	131
KOR	0.8095	0.9141	45.4	0.8298	0.9323	7.0	143
JPN	0.8090	0.8789	44.9	0.8221	0.8831	17.0	125
ISL	0.8030	0.9315	5.4	0.7884	0.9446	3.0	102
NZL	0.8022	0.8924	13.0	0.8026	0.8889	6.0	142
GBR	0.7994	0.8841	36.4	0.7991	0.8872	10.0	287
AUS	0.7982	0.8698	45.4	0.7942	0.8635	23.0	200
IRL	0.7919	0.8888	23.7	0.8052	0.8913	11.0	136
SWE	0.7853	0.9033	10.8	0.7823	0.9073	5.0	127
BEL	0.7837	0.8223	182.7	0.8107	0.8419	53.5	178
CZE	0.7789	0.8886	13.2	0.7725	0.8903	5.0	210
NOR	0.7736	0.8660	18.4	0.7762	0.8637	7.0	135
CHL	0.7731	0.8769	5.9	0.7612	0.8791	3.0	125
CHE	0.7711	0.8682	24.7	0.7657	0.8726	6.0	223
FRA	0.7699	0.8393	55.6	0.7711	0.8479	22.0	122
DEU	0.7663	0.8962	11.9	0.7741	0.9425	3.0	187
DNK	0.7658	0.8433	53.4	0.7643	0.8489	18.5	146
USA	0.7626	0.8644	34.8	0.7674	0.8642	10.0	98
BRA	0.7616	0.8560	62.3	0.7481	0.8591	6.0	195
ESP	0.7493	0.8247	157.4	0.7449	0.7953	85.0	153
LIE	0.7463	0.8552	49.8	0.7621	0.8846	12.0	9
ITA	0.7362	0.7906	85.3	0.7456	0.7940	32.5	134
ISR	0.7234	0.8181	29.4	0.7389	0.8152	7.0	45
PRT	0.7230	0.7885	72.7	0.7433	0.7914	36.0	143
RUS	0.7219	0.8199	62.4	0.7190	0.8206	24.0	215
THA	0.7213	0.8457	49.6	0.7117	0.8489	12.0	165
LUX	0.7153	0.7956	35.2	0.6930	0.7907	6.0	19
POL	0.7111	0.7805	272.8	0.7111	0.7708	68.0	113
GRC	0.7049	0.7790	68.1	0.7293	0.7955	46.5	154
LVA	0.6944	0.7752	125.8	0.6928	0.7828	55.0	107
ROM	0.6791	0.7648	174.5	0.7001	0.7827	77.0	171
MKD	0.6768	0.7679	35.4	0.6702	0.7636	9.5	60
BGR	0.6435	0.7215	149.3	0.6422	0.7201	83.0	120
ALB	0.6347	0.7390	46.5	0.6347	0.7281	9.0	120
ARG	0.6314	0.6890	401.9	0.6285	0.6891	245.5	118
PER	0.6198	0.7253	55.4	0.6072	0.7079	22.0	99
IDN	0.6184	0.7129	81.0	0.6143	0.7125	25.5	232
TOTAL	0.7485	0.8384	67.6	0.7563	0.8503	13.0	5528

Table 7: Efficiency estimates of countries' hypothetical, median schools, input orientation, $p = q = 3$, with bootstrap 95-percent confidence interval estimates.

	$\widehat{\mathcal{I}}$	$\widehat{\text{BIAS}}$	std. dev.	$\widehat{\widehat{\mathcal{I}}}$	$\widehat{\mathcal{I}}_{lo}$	$\widehat{\mathcal{I}}_{hi}$
GBR	1.3087	-0.0738	0.02688	1.3826	1.3442	1.4316
NZL	1.4920	-0.0945	0.03456	1.5864	1.5366	1.6479
BRA	1.5886	-0.047	0.01917	1.6357	1.6078	1.6700
ISL	1.6171	-0.0943	0.03433	1.7113	1.6592	1.7718
CHL	1.6052	-0.1353	0.04992	1.7405	1.6708	1.8325
NOR	1.7623	-0.0716	0.02978	1.8340	1.7915	1.8887
MEX	1.7455	-0.1278	0.03266	1.8733	1.8197	1.9276
DEU	1.7945	-0.060	0.0155	1.8546	1.8289	1.8803
NLD	1.7642	-0.1145	0.02977	1.8787	1.8294	1.9274
KOR	1.8446	-0.0846	0.03401	1.9293	1.8792	1.9890
AUS	1.8722	-0.0444	0.01344	1.9165	1.8957	1.9404
HKG	1.8867	-0.0455	0.01836	1.9322	1.9071	1.9671
USA	1.9141	-0.0614	0.02091	1.9755	1.9448	2.0124
JPN	1.9308	-0.0465	0.01366	1.9773	1.9578	2.0022
SWE	1.9570	-0.0631	0.02364	2.0200	1.9863	2.0638
LIE	1.9206	-0.1329	0.04203	2.0535	1.9917	2.1271
FIN	2.0479	-0.0909	0.0296	2.1388	2.0935	2.1904
ALB	2.0403	-0.1472	0.06054	2.1875	2.0989	2.2957
IRL	2.0933	-0.0624	0.01847	2.1557	2.1272	2.1877
ISR	2.0816	-0.1127	0.03997	2.1943	2.1307	2.2618
CHE	2.1222	-0.0698	0.02372	2.1921	2.1555	2.2317
MKD	2.1292	-0.1252	0.04697	2.2544	2.1800	2.3361
BEL	2.1557	-0.0637	0.02246	2.2194	2.1854	2.2589
PER	2.1577	-0.0685	0.02237	2.2262	2.1913	2.2642
ESP	2.1669	-0.0551	0.01725	2.2220	2.1944	2.2520
FRA	2.1751	-0.0615	0.01834	2.2367	2.2080	2.2692
DNK	2.2156	-0.0604	0.02009	2.2759	2.2455	2.3119
CZE	2.2236	-0.0896	0.03315	2.3132	2.2603	2.3703
ITA	2.3707	-0.1046	0.03832	2.4753	2.4154	2.5428
THA	2.4871	-0.0814	0.03002	2.5685	2.5242	2.6218
IDN	2.5294	-0.1668	0.04207	2.6962	2.6165	2.7611
GRC	2.6313	-0.1365	0.04861	2.7677	2.6919	2.8484
RUS	2.8082	-0.1713	0.05355	2.9795	2.8945	3.0705
ROM	2.8691	-0.1008	0.0361	2.9699	2.9148	3.0330
LUX	2.7308	-0.4398	0.1393	3.1706	2.9345	3.3903
BGR	3.0661	-0.0749	0.02638	3.1410	3.0998	3.1851
POL	3.1305	-0.1072	0.03211	3.2377	3.1879	3.2924
LVA	3.2232	-0.1381	0.04148	3.3614	3.2940	3.4304
PRT	3.3301	-0.4153	0.1429	3.7454	3.5009	3.9712
ARG	3.4709	-0.1175	0.04378	3.5884	3.5240	3.6640

Table 8: Efficiency estimates of countries' median schools, output orientation, $p = q = 3$.

	$\widehat{\mathcal{O}}$	$\widehat{\text{BIAS}}$	std. dev.	$\widehat{\widehat{\mathcal{O}}}$	$\widehat{\mathcal{O}}_{lo}$	$\widehat{\mathcal{O}}_{hi}$
GBR	1.0000	1e-040	9.633e-05	0.9999	0.9998	1.0001
NZL	1.0000	1e-040	9.287e-05	0.9999	0.9997	1.0000
CHL	0.9252	0.0057	0.002035	0.9195	0.9158	0.9222
KOR	0.9148	0.0015	0.0005561	0.9134	0.9124	0.9142
HKG	0.9102	0.0015	0.000485	0.9088	0.9079	0.9095
ARG	0.9091	0.0000	2.663e-05	0.9091	0.9091	0.9092
ESP	0.9091	0.0000	5.616e-05	0.9091	0.9090	0.9092
FRA	0.9091	0.0000	5.613e-05	0.9091	0.9090	0.9092
GRC	0.9091	1e-040	0.0001165	0.9090	0.9088	0.9092
BEL	0.9091	0.0000	9.606e-05	0.9090	0.9089	0.9092
ITA	0.9091	0.0000	6.151e-05	0.9091	0.9090	0.9092
RUS	0.9091	0.0000	6.483e-05	0.9091	0.9089	0.9092
AUS	0.9091	1e-040	0.0001331	0.9090	0.9087	0.9091
USA	0.9091	2e-040	0.0001941	0.9089	0.9085	0.9091
CZE	0.9091	2e-040	0.0001784	0.9089	0.9085	0.9091
ISR	0.9091	3e-040	0.0002699	0.9088	0.9082	0.9091
JPN	0.9091	4e-040	0.0002651	0.9087	0.9082	0.9091
NOR	0.9091	0.0013	0.0008602	0.9078	0.9061	0.9089
ALB	0.9104	0.0035	0.00136	0.9068	0.9046	0.9087
ISL	0.9091	0.0102	0.003495	0.8989	0.8937	0.9049
NLD	0.8630	0.0044	0.001668	0.8586	0.8557	0.8609
LIE	0.8448	0.0022	0.001282	0.8426	0.8403	0.8442
FIN	0.8390	0.0046	0.001455	0.8343	0.8321	0.8367
IRL	0.8316	0.0025	0.001109	0.8291	0.8272	0.8309
DEU	0.8277	0.0038	0.001226	0.8239	0.8218	0.8257
MEX	0.8255	0.0087	0.004138	0.8168	0.8104	0.8226
MKD	0.8237	0.0023	0.0008252	0.8214	0.8200	0.8225
PER	0.8210	0.0010	0.0004728	0.8200	0.8191	0.8207
SWE	0.8230	0.0054	0.001477	0.8177	0.8152	0.8200
THA	0.8202	0.0016	0.0005306	0.8186	0.8177	0.8195
BGR	0.8182	0.0000	3.441e-05	0.8182	0.8181	0.8183
POL	0.8182	0.0000	5.935e-05	0.8182	0.8181	0.8182
LVA	0.8182	0.0000	4.963e-05	0.8182	0.8181	0.8182
PRT	0.8182	2e-040	0.0001341	0.8180	0.8178	0.8182
ROM	0.8182	4e-040	0.0002493	0.8178	0.8173	0.8181
LUX	0.8182	9e-040	0.0005513	0.8173	0.8163	0.8180
IDN	0.8182	0.0012	0.0004785	0.8170	0.8162	0.8178
DNK	0.8182	0.0020	0.0009351	0.8162	0.8146	0.8177
CHE	0.8195	0.0034	0.0009624	0.8161	0.8145	0.8176
BRA	0.7547	0.0049	0.002482	0.7498	0.7450	0.7528

Table 9: Percentage of countries' schools projected onto increasing, constant, and decreasing returns-to-scale regions of the production frontier ($p = q = 3$).

	Input orientation			Output orientation		
	IRS	CRS	DRS	IRS	CRS	DRS
ALB	12.50	65.00	22.50	0.00	0.00	100.00
ARG	33.05	4.24	62.71	0.00	0.00	100.00
AUS	0.50	1.00	98.50	0.00	0.00	100.00
BEL	24.16	1.69	74.16	0.00	0.56	99.44
BGR	79.17	0.00	20.83	0.00	0.00	100.00
BRA	78.97	3.59	17.44	5.64	5.13	89.23
CHE	31.39	2.24	66.37	0.00	2.24	97.76
CHL	36.80	30.40	32.80	2.40	7.20	90.40
CZE	11.90	4.76	83.33	0.00	0.00	100.00
DEU	40.64	0.53	58.82	0.00	0.53	99.47
DNK	28.08	2.05	69.86	0.68	1.37	97.95
ESP	9.15	0.00	90.85	0.00	0.00	100.00
FIN	2.17	0.00	97.83	0.00	0.00	100.00
FRA	36.07	0.82	63.11	0.00	0.00	100.00
GBR	0.00	2.09	97.91	0.00	1.05	98.95
GRC	8.44	6.49	85.06	0.00	0.00	100.00
HKG	4.55	1.52	93.94	0.00	0.00	100.00
IDN	68.53	19.83	11.64	0.00	0.86	99.14
IRL	2.21	0.74	97.06	0.00	0.00	100.00
ISL	0.00	1.96	98.04	0.00	0.98	99.02
ISR	11.11	20.00	68.89	0.00	2.22	97.78
ITA	8.21	2.24	89.55	0.00	0.75	99.25
JPN	0.00	0.80	99.20	0.00	0.00	100.00
KOR	2.80	2.10	95.10	0.00	0.00	100.00
LIE	22.22	0.00	77.78	0.00	0.00	100.00
LUX	73.68	0.00	26.32	0.00	0.00	100.00
LVA	49.53	9.35	41.12	0.00	0.00	100.00
MEX	39.69	29.77	30.53	2.29	9.16	88.55
MKD	85.00	5.00	10.00	0.00	0.00	100.00
NLD	18.84	1.45	79.71	0.00	1.45	98.55
NOR	0.00	0.74	99.26	0.00	0.00	100.00
NZL	0.00	0.00	100.00	0.00	0.00	100.00
PER	50.51	10.10	39.39	0.00	1.01	98.99
POL	55.75	0.00	44.25	0.00	0.00	100.00
PRT	48.95	0.00	51.05	0.00	0.00	100.00
ROM	77.19	1.17	21.64	0.00	0.00	100.00
RUS	15.81	24.19	60.00	0.00	0.00	100.00
SWE	11.02	0.00	88.98	0.00	0.00	100.00
THA	52.73	6.67	40.61	0.00	0.61	99.39
USA	18.37	1.02	80.61	0.00	1.02	98.98
TOTAL	27.50	6.62	65.88	0.33	0.94	98.73

Table 10: Mean input, output levels for 10-percent input-efficient, 10-percent input-inefficient schools ($p = q = 3$).

	#	mean	std. dev.
	Input Eff. > 2.0259		
social	837	3.55	0.85
teach	837	47.72	27.59
propcert	837	0.83	0.26
test	837	911.50	139.04
grade	837	9.27	0.58
students	837	624.33	417.67
	Input Eff. ≤ 2.0259		
social	279	3.07	0.87
teach	279	34.78	21.90
propcert	279	0.53	0.43
test	279	949.94	187.50
grade	279	9.55	0.58
students	279	690.27	489.89

Table 11: Mean input, output levels for 10-percent output-efficient, 10-percent output-inefficient schools ($p = q = 3$).

	#	mean	std. dev.
	Output Eff. < 0.9090		
social	669	3.48	0.96
teach	669	49.16	27.74
propcert	669	0.82	0.29
test	669	906.47	146.91
grade	669	8.90	0.30
students	669	674.46	450.51
	Output Eff. ≥ 0.9090		
social	447	3.35	0.75
teach	447	37.49	23.88
propcert	447	0.66	0.38
test	447	943.02	160.36
grade	447	9.99	0.13
students	447	590.47	412.80

Table 12: Potential percentage reductions in inputs, and potential percentage increases in outputs, from elimination of technical inefficiency ($p = q = 3$).

	Potential input reductions			Potential Output increases		
	social	**teach**	**propcert**	**test**	**grade**	**students**
ALB	49.21	49.76	56.23	10.57	10.59	10.27
ARG	68.69	68.65	67.38	14.48	15.02	15.77
AUS	44.81	39.79	41.52	8.20	8.24	8.08
BEL	53.22	53.38	53.48	11.50	11.99	11.34
BGR	64.88	64.03	64.96	22.23	22.40	22.41
BRA	32.98	34.32	33.19	22.84	22.92	19.32
CHE	46.86	50.35	46.96	17.29	17.66	17.87
CHL	31.58	31.44	36.44	11.72	12.14	10.67
CZE	51.16	51.39	50.68	13.33	13.38	14.16
DEU	41.94	38.71	40.09	16.79	17.42	15.76
DNK	49.49	49.22	52.51	19.26	19.41	19.65
ESP	51.84	51.60	49.74	10.83	10.85	10.98
FIN	48.75	49.56	48.93	17.57	17.73	18.10
FRA	51.24	48.48	53.08	14.66	14.99	13.42
GBR	25.44	22.33	25.11	2.11	2.02	1.97
GRC	59.05	58.85	57.89	10.93	11.02	10.99
HKG	47.17	45.58	46.12	9.53	9.67	9.54
IDN	54.62	56.44	57.73	18.81	18.60	17.58
IRL	49.83	48.22	48.94	14.24	14.07	14.20
ISL	34.55	38.74	34.69	8.02	8.15	9.29
ISR	46.78	42.01	49.49	10.33	10.21	9.42
ITA	54.63	54.80	54.69	10.64	10.83	10.80
JPN	47.54	46.54	46.74	9.18	9.27	9.11
KOR	46.54	44.70	45.54	9.43	9.47	9.06
LIE	51.18	50.23	50.37	18.65	18.88	17.76
LUX	55.91	54.01	59.06	20.68	20.82	20.44
LVA	63.03	62.53	64.44	18.40	18.41	17.04
MEX	29.33	27.19	30.54	11.20	11.48	7.99
MKD	49.77	47.59	56.14	19.10	19.18	18.94
NLD	41.54	40.31	39.71	12.59	13.00	11.95
NOR	42.37	42.40	40.80	8.96	9.01	9.36
NZL	29.44	26.91	27.53	0.00	0.00	0.00
PER	48.95	48.24	49.30	15.20	15.62	13.25
POL	62.68	63.37	62.36	19.60	20.03	19.58
PRT	61.75	58.66	61.19	17.89	18.08	16.71
ROM	64.44	64.76	63.97	23.54	23.88	22.85
RUS	59.22	59.35	60.79	11.98	11.93	10.99
SWE	47.66	48.65	46.97	19.21	19.37	19.54
THA	49.63	44.80	48.44	16.29	16.33	13.34
USA	48.58	44.53	45.11	12.44	12.53	11.82

Figure 1: Kernel density estimates for input and output variables.

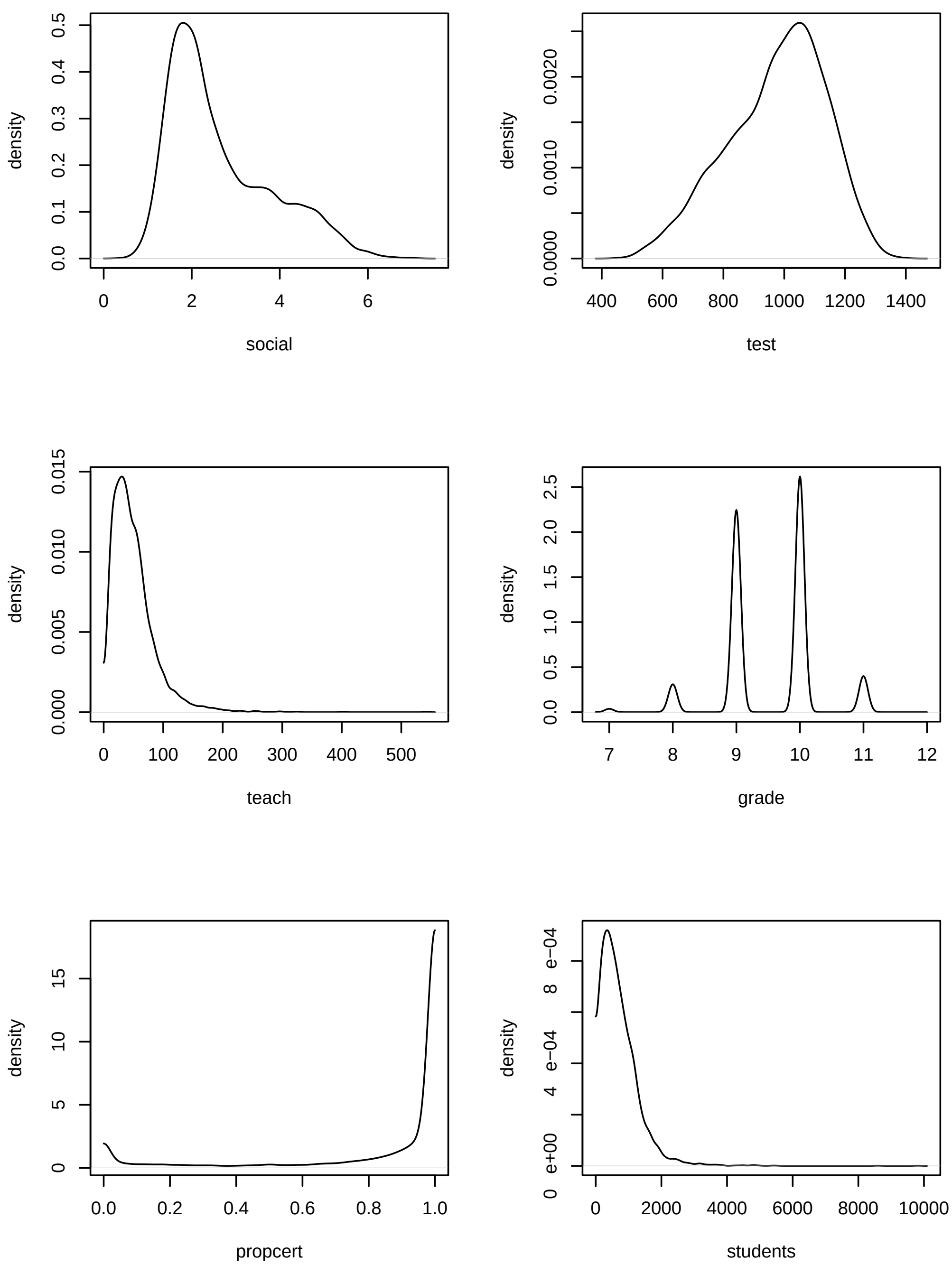

Figure 2: 95-percent confidence interval estimates, input efficiency, transitioning countries, $p = q = 3$.

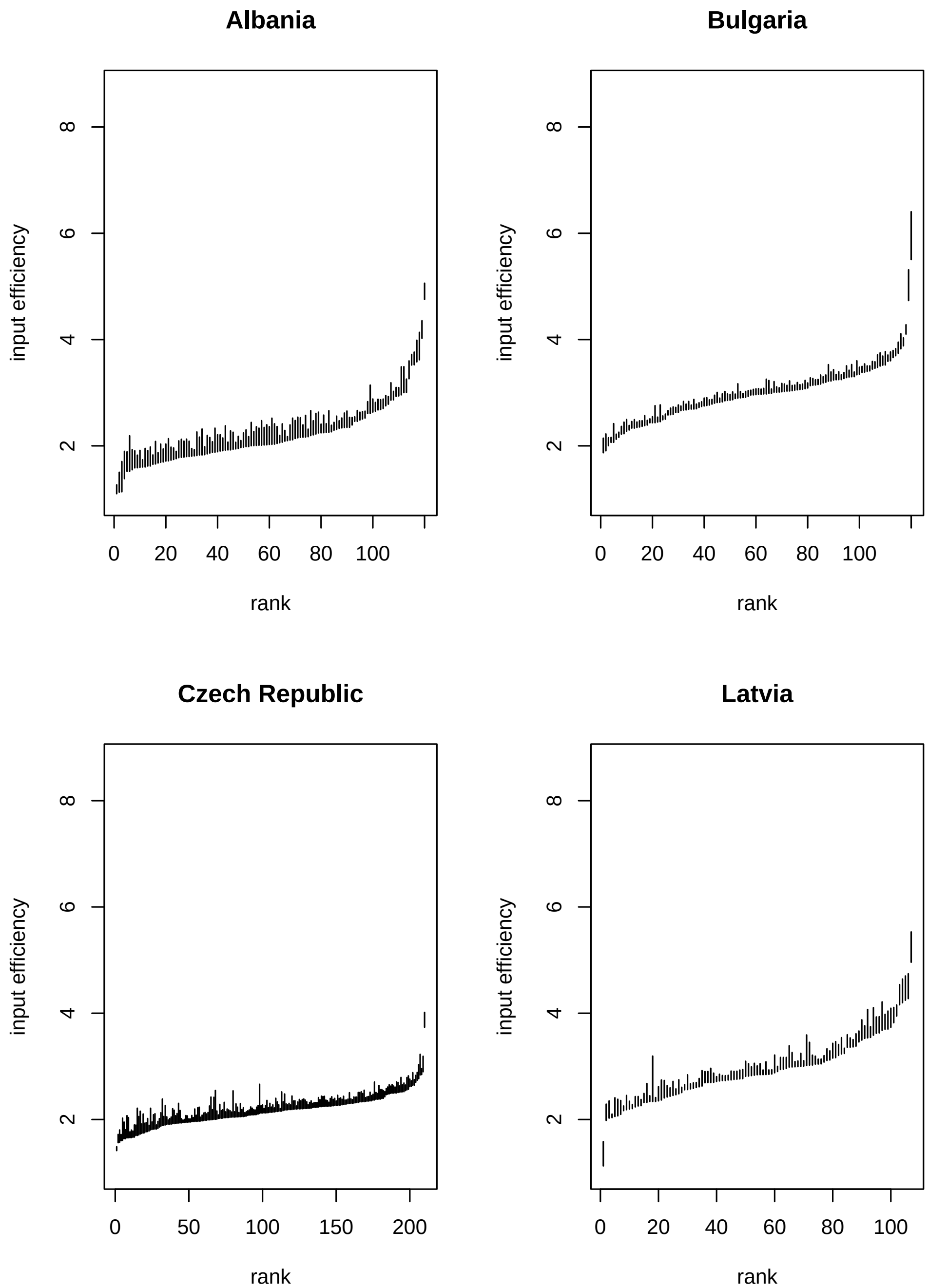

Figure 3: 95-percent confidence interval estimates, input efficiency, transitioning countries, $p = q = 3$.

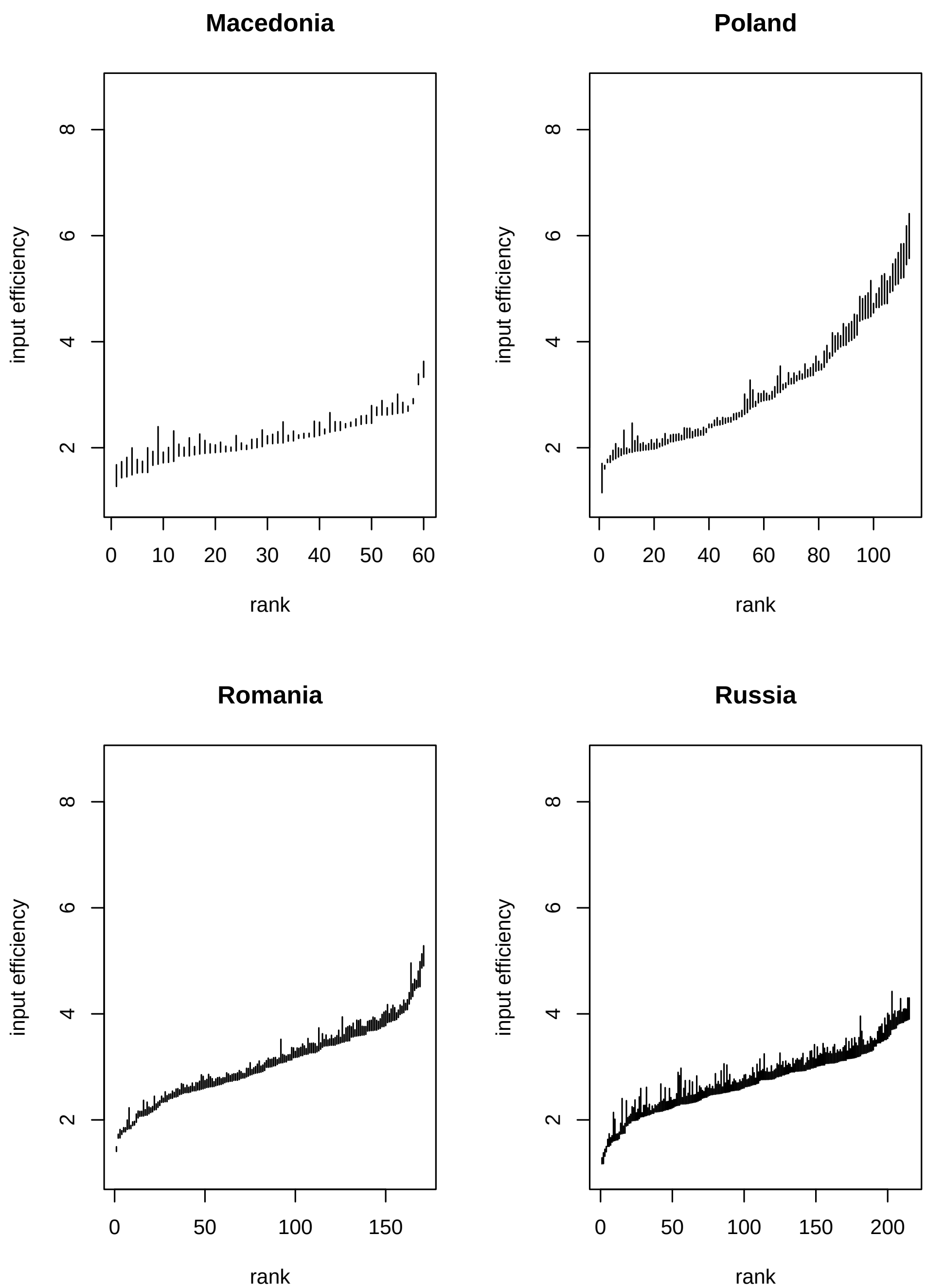

Figure 4: 95-percent confidence interval estimates, output efficiency, transitioning countries, $p = q = 3$.

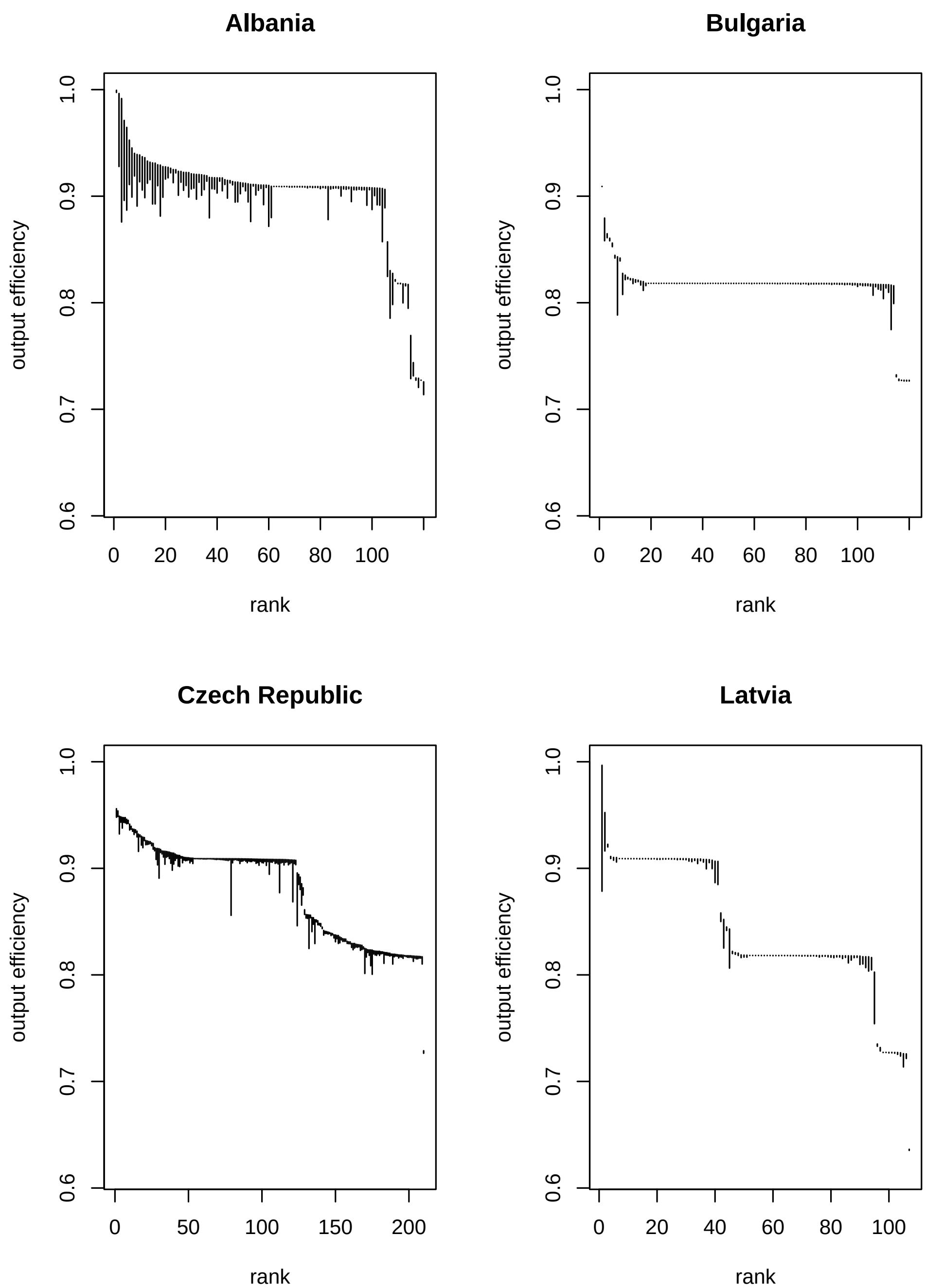

Figure 5: 95-percent confidence interval estimates, output efficiency, transitioning countries, $p = q = 3$.