МАКРО- И МИКРОЭКОНОМИЧЕСКИЕ ИССЛЕДОВАНИЯ

JEL: C61, C63

Estimation of Cost Efficiency in Non-parametric Frontier Models

G. Besstremyannaya[1], *J. Simm*[2]

[1] New Economic School and Institute of the Far East Studies of the Russian Academy of Sciences, 45, Skolkovskoe shosse, Moscow, 121353, Russian Federation
[2] Stadius Centre for Dynamical Systems, Signal Processing and Data Analytics, Belgium, B–3001, Heverlee

For citation: Besstremyannaya, G., Simm, J. (2019) Estimation of Cost Efficiency in Non-parametric Frontier Models. *St Petersburg University Journal of Economic Studies*, vol. 35, issue 1, pp. 3–25. https://doi.org/10.21638/spbu05.2019.101

The paper proposes a bootstrap methodology for estimating cost efficiency in data envelopment analysis. We consider the conventional concept of Fare, Grosskopf and Lovellcost efficiency, for which our algorithm re-samples "naive" input-oriented efficiency scores, rescales original inputs to bring them to the frontier, and then re-estimates cost efficiency scores for the rescaled inputs. Next, we examine Tone cost efficiency, where input prices vary across producers. Here we show that the direct modification on bootstrap algorithms by Simar and Wilson are applicable. We consider cases both with the absence and presence of environmental variables (i.e. input variables not directly controlled by firms). The bootstrap methodology exploits these assumptions: 1) the sample are i.i.d. random variables with the continuous joint probability density function with support over production set; 2) the frontier is smooth; and 3) the probability of observing firms on the frontier approaches unity with an increase in sample. The results of simulations for a multi-input, multi-output Cobb–Douglas production function with correlated outputs, and correlated technical and cost efficiency, show consistency of our proposed algorithm, even for small samples. Finally, we offer real data estimates for the Japanese banking industry in 2013. Our package "rDEA," developed in the R language, is available from the GitHub and CRAN repository.

Keywords: bootstrap, price efficiency, border analysis, banking.

Introduction

Data envelopment analysis (*DEA*) was introduced in [Charnes, Cooper, Rhodes, 1978] as a linear optimization technique, stemming from the seminal work of [Farrell, 1957], who defined a firm's technical and price efficiency and proposed a method for constructing a frontier as a linear convex hull surface to envelop observations. The efficiency scores of each firm are estimated according to the distance from the empirical frontier, which is treated as fully efficient firms. The analysis is based on common premises about production: monotonicity of technology, requirement of inputs for production ("no free lunch" condition), closedness, and strict convexity of the production set [Simar, Wilson, 2000b]. However, the empirical frontier may fail to incorporate unobservable but very efficient firms [Simar, Wilson, 1998]. Thus, the efficiency scores, linked to the empirical frontier, are upwardly biased. Standard approaches for consistent correction of bias in the case of technical efficiency scores are a homogeneous bootstrap based on re-sampling from a smooth consistent estimator of the joint density of input-output pairs, or a semi-parametric bootstrap in the presence of additional inputs, i.e. so-called environmental variables, which are not directly controlled by producers [Simar, Wilson, 2000b; Simar, Wilson, 1998; 2007][1]. The bootstrap methodology exploits these assumptions about the data-generating process for observations of firms, which enable a consistent approximation of the unknown distribution of efficiency scores: 1) the sample is i.i.d. random variables with a continuous joint probability density function with support over the production set; 2) the frontier is smooth; and 3) the probability of observing firms on the frontier approaches unity with an increase in sample size (see [Kneip, Simar, Wilson, 2008; Simar, Wilson, 2000b]).

Concerning cost minimization DEA, as formulated in [Fare, Grosskopf, Lovell, 1985, practitioners suggest a direct modification of the [Simar, Wilson, 1998; 2007] bootstrap (e.g. [Borger, Kerstens, Staat, 2008]).

In this paper we show that a direct modification the [Simar, Wilson,1998; 2007] bootstrap is inconsistent for cost minimization DEA, and we propose an alternative algorithm. The proposed algorithm re-samples "naive" input-oriented efficiency scores, rescales original inputs to bring them to the frontier, and then re-estimates cost efficiency scores for the rescaled inputs. The results of the simulations for a multi-input, multi-output Cobb–Douglas production function with correlated outputs, and correlated technical and cost efficiency, show consistency of our proposed algorithm, even for small samples. As for a recently defined "new" cost efficiency [Tone, 2002], which to the best of our knowledge is commonly assessed only in terms of naïve scores, we demonstrate that the direct modification of the [Simar, Wilson, 1998; 2007] bootstrap is consistent. Finally, we apply the algorithm to real data from 106 Japanese banks for fiscal year 2013.

The remainder of the paper is structured as follows. Section 1 reviews the theoretical framework for bias correction of technical efficiency scores, using an example of input orientation. Section 2 demonstrates the inconsistency of a direct application of [Simar, Wilson, 1998; 2007] bootstrap and offers an alternative bootstrap algorithm for robust estimation of [Fare, Grosskopf, Lovell, 1985] cost efficiency in absence and in presence

[1] In the absence of environmental variables, the smooth bootstrap provides a better inference in a non-simulation context [Kneip, Simar, Wilson, 2008] than an alternative bootstrap based on subsampling [Simar, Wilson, 2011a].

of environmental variables. Section 3conducts simulations of production frontier and technical and cost inefficiencies. Section 4provides real data estimates with a nationwide sample of Japanese banks. The Appendix sets up the microeconomic framework for estimating technical and cost inefficiencies. Our computations are conducted with an R package "rDEA" [Simm, Besstremyannaya, 2016], which is available from GitHub and CRAN repositary.

1. Estimates of input-oriented efficiency

1.1. Naive score

Denote the estimator of existing technology (of the production set), which produces outputs $\boldsymbol{y}_m (m=1,\ldots,M)$ using inputs $\boldsymbol{x}_n (n=1,\ldots,N)$ as $\hat{T} = \{(\boldsymbol{x},\boldsymbol{y}): \boldsymbol{x} \geq \boldsymbol{X\lambda}, \boldsymbol{y} \leq \boldsymbol{Y\lambda}, \boldsymbol{\lambda} \geq \boldsymbol{0}\}$ where efficiency for a given unit $(\mathbf{x}, \mathbf{y})$ is measured relative to the boundary of the convex hull $(\mathbf{X}, \mathbf{Y})$ under the assumption of constant returns to scale. Input set $L(\mathbf{y})$ [Coelli, Rao, Battese,1994; Shephard,1981] contains inputs that can produce a given amount of output, so that $L(\mathbf{y}) = \{(\mathbf{x}) : (\mathbf{x}, \mathbf{y}) \in T\}$. The technology satisfies common premises about inputs, outputs, and the production set, which were briefly outlined in the Introduction. The important assumptions about the production process are strict convexity of $L(\mathbf{y})$ and strong (free) disposability of inputs and outputs. In particular, strong disposability of inputs implies that if $\mathbf{x} \in L(\mathbf{y})$, and if $\mathbf{x} \geq \mathbf{x}$, then $\mathbf{x} \in L(\mathbf{y})$. The input-oriented efficiency θ_j for a given firm j $(j=1,\ldots,J)$ is defined as a solution to the optimization problem below (constant returns to scale, *CRS* formulation of [Charnes, Cooper, Rhodes, 1978]:

$$\min_{\theta_j, \lambda} \theta_j$$

$$s.t. -y_{mj} + \sum_{i=1}^{J} \lambda_i y_{mi} \geq 0, \;\; m = 1,\ldots,M, \tag{1}$$

$$\theta_j x_{nj} - \sum_{i=1}^{J} \lambda_i x_{ni} \geq 0, \;\; n = 1,\ldots,N,$$

$$\lambda_i \geq 0, \;\; i = 1,\ldots,J.$$

Additional constraints $\sum_{i=1}^{J} \lambda_i x_{ni} = 1$ impose variable returns to scale (*VRS*).

It should be noted that system (1) represents a linear maximization program written in concise notation. In fact, for each firm j there is a set of M constraints, where each constraint corresponds to a particular output y_{mj}. Similarly, there are N constraints on each input x_{nj}.

The economic intuition behind the system (1) may be explained as follows. We examine the behavior of economic agents and construct the production possibility set, using inputs and outputs for each agent. The analysis focuses on minimizing the amount of inputs required to produce a given amount of output; therefore, the problem is called "input-oriented *DEA*". The boundary of the production possibility set becomes the production possibility frontier. Thus, the unity values of θ_j imply that the agent is fully efficient, as these points lie on the non-linear hull of observations (i. e. on the empirical estimate of the production possibility frontier). The weights λ_i in system (1) are exploited for the construction of the frontier.

Another optimization problem, called "output-oriented *DEA*," studies the maximization of the output under a given amount of inputs. The reciprocals of efficiency scores become the multipliers of the output coordinates:

$$\min_{\theta_j, \lambda} \theta_j$$

$$s.t. -y_{mj} / \theta_j + \sum_{i=1}^{J} \lambda_i y_{mi} \geq 0, \ \ m = 1, \ldots, M, \tag{1a}$$

$$x_{nj} - \sum_{i=1}^{J} \lambda_i x_{ni} \geq 0, \ \ n = 1, \ldots, N,$$

$$\lambda_i \geq 0, \ \ i = 1, \ldots, J.$$

Output-oriented *DEA* may be used to analyze the behavior of such economic agents as municipal hospitals, which have difficulties increasing the amount of inputs (medical personnel) and hence are analyzed in terms of their ability to maximize outputs (number of treated patients, see [Besstremyannaya, 2013]).

In this paper we focus on an input-oriented *DEA* when proposing our estimations of bias-corrected cost efficiency. Note, however, that the bottstrap algorithm may be straightforwardly modified for using the output-oriented score at the first stages.

1.2. Bias correction of naive DEA score

The estimates of input-oriented efficiency are upwards biased, since the estimated boundary $\hat{L}^{\partial}(\mathbf{y})$ of the input set is based on the sample of observed units, which may fail to incorporate the most efficient units in the true $L(\mathbf{y})$ [Simar, Wilson 1998; 2000]. The bootstrap methods were proposed in [Simar, Wilson, 1998] and the assumptions about the data-generating process are further explained in [Simar, Wilson, 2000b; Kneip, Simar, Wilson, 2008].The bootstrap corrects for the bias by constructing pseudo-samples that would belong to $\hat{L}(\mathbf{y})$. The central idea for using the bootstrap is an assumption that the empirical bootstrap distribution of the DEA efficiency score consistently estimates the sampling distribution, and the assumption is ensured with the use of smoothing techniques [Simar, Wilson, 2000a]. Then, according to the re-centering idea of bootstrap, $\text{bias}\hat{\theta}_i = E(\hat{\theta}_i) - \theta_i$. So the estimator of the bias becomes $\text{bias}\hat{\theta}_i = \text{bias}\hat{\theta}_i^* = E(\hat{\theta}_i^*) - \hat{\theta}_i$. In particular, the homogeneous smoothed bootstrap projects each observation on the frontier and then "pushes" it inside the $\hat{L}(\mathbf{y})$ [Simar, Wilson, 2008; 1998].

1. Estimate naive scores $\hat{\theta}_1, \ldots, \hat{\theta}_J$, for each $i = 1, \ldots, J$ according to system (1). Assume $(\theta_1, \ldots \theta_J)$ are i.i.d. with pdf $f(\cdot)$.
2. Loop B times to obtain J sets of bootstrap estimates $\{\hat{\theta}_{ib}^*\}_{b=1}^{B}$.

1. Obtain a smooth estimate $\hat{f}(\theta)$ and for each $i=1,...,J$ draw θ^{*}_{ib} from this estimate[2].
2. Assume homogeneous distribution of joint density of θ in input-output space, i.e. $\hat{f}(\theta_i \mid (\mathbf{x}_i, \mathbf{y}_i) = \hat{f}(\theta_i)$ and assign $\boldsymbol{x}^{*}_{ib} = \frac{\hat{\theta}_i}{\theta^{*}_{ib}} \boldsymbol{x}_i$.
3. Calculate $\hat{\theta}^{*}_{ib}$ for $(\mathbf{x}^{*}_{ib}, \mathbf{y}_i)$.

3. $\text{bias}\hat{\hat{\theta}}_i = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^{*}_{ib} - \hat{\theta}_i$ and bias-corrected score $\hat{\hat{\delta}}_i = \hat{\delta}_i - \text{bias}\hat{\hat{\delta}}_i$.

Rescaling at step (2.2) guarantees that pseudo-samples $\{(\mathbf{x}^{*}_{ib}, \mathbf{y}_i)\}^{B}_{b=1} \in \hat{L}(\mathbf{y})$. Indeed, input-oriented efficiency evaluates the potential of DMUi for maximal reduction of inputs, holding the amount of outputs constant. The constraints $xi \geq X\lambda$ imply inputs are larger than possible. Therefore, multiplication of each input by $\hat{\theta}_i$, $0 \leq \hat{\theta}_i \leq 1$, projects it to $\hat{L}^{\partial}(\mathbf{y})$, so that the projected observation become an estimate of an efficient input level with coordinates $(\hat{\theta}_i \mathbf{x}_i, \mathbf{y}_i)$. The assumption about homogeneous distribution of joint density of θ allows drawing each θ^{*}_{ib} for pseudo-samples from the same estimate of $\hat{f}(\theta)$, which is obtained for the original sample. Therefore, division of each projected input by $\theta^{*}_{ib}, 0 \leq \theta^{*}_{ib} \leq 1$ in step (2.2) "pushes" the projected input inside $\hat{L}(\mathbf{y})$.

In presence of an r-dimensional vector of environmental variables z, i.e. a special type of inputs that are not directly controlled by producers, [Simar, Wilson, 2007] propose a semi-parametric bootstrap for correcting the bias of the distance function score δ, the reciprocal of θ[3]. The algorithm, in the case of input-orientation, is based on the premise about the separability of inputs and environmental variables, i.e. the fact that the support of x does not depend on z [Simar, Wilson, 2011b].

1. Estimate the naïve distance function scores $\hat{\delta}_1, ..., \hat{\delta}_J$, for each $i=1,...,J$ using the equivalent of system (2) for reciprocals of θ. Assume, $\delta_i = \mathbf{z}_i \hat{\boldsymbol{a}} + \varepsilon_i \geq 1$ where ε_i are i. i. d. and independent from zi, $\varepsilon_i \sim N(0, \sigma^2_{\varepsilon})$ with left truncation at $(1 - \mathbf{z}_i \hat{\boldsymbol{a}})$.
2. Use observations for which $\hat{\delta} > 1$ to obtain $\hat{\boldsymbol{a}}$ and $\hat{\sigma}_{\varepsilon}$ in the truncated regression $\hat{\delta}_i = \mathbf{z}_i \hat{\boldsymbol{a}} + \varepsilon_i \geq 1$.
3. Loop *B* times to obtain *J* sets of bootstrap estimates $\{\hat{\delta}^{*}_{ib}\}^{B}_{b=1}$.
   1. For each $i=1,...,J$ draw ε_i from $N(0, \hat{\sigma}^2_{\varepsilon})$ with left truncation at $(1 - \mathbf{z}_i \hat{\boldsymbol{a}})$.
   2. For each $i=1,...,J$ compute $(\delta^{*}_i = \mathbf{z}_i \hat{\boldsymbol{a}} + \varepsilon_i)$.
   3. Assign $\boldsymbol{x}^{*}_{ib} = \frac{\delta^{*}_{ib}}{\hat{\delta}_i} \boldsymbol{x}_i$.
   4. Calculate $\hat{\delta}^{*}_{ib}$ for $(\mathbf{x}^{*}_{ib}, \mathbf{y}_i)$.
4. $\text{bias}\hat{\delta}_i = \frac{1}{B}\sum_{b=1}^{B}\hat{\delta}^{*}_{ib} - \hat{\delta}_i$ and bias-corrected score $\hat{\hat{\delta}}_i = \hat{\delta}_i - \text{bias}\hat{\hat{\delta}}_i$.

---

[2] Smoothing is necessary to avoid inconsistency in estimating the upper bound of the support of the underlying data-generating process $f(\cdot)$ [Simar, Wilson, 1998].

[3] θ, which is bounded between 0 and 1, could not be used to estimatethe truncated regression [Simar, Wilson, 2008].

2. Estimates of cost efficiency

2.1. Naive score with given input prices

Denote $\mathbf{w}_j$ the vector of input prices. Fare, Grosskopf, and Lovell (1985) define cost efficiency γ_j as

$$\gamma_j = \mathbf{w}_j \mathbf{x}_j^{opt} / \mathbf{w}_j \mathbf{x}_j, \tag{2}$$

where $\mathbf{x}_j^{opt}$ is a solution to the optimization problem (formulated below for constant returns to scale):

$$\begin{gathered} \min_{\boldsymbol{x}_j, \lambda} \boldsymbol{w}_j \boldsymbol{x}_j \\ s.t. -y_{mj} + \sum_{i=1}^{J} \lambda_i y_{mi} \geq 0, \ \ m = 1, \ldots, M, \\ x_{nj} - \sum_{i=1}^{J} \lambda_i x_{ni} \geq 0, \ \ n = 1, \ldots, N, \\ \lambda_i \geq 0, \ \ i = 1, \ldots, J. \end{gathered} \tag{3}$$

According to (2) and system (3), $0 \leq \gamma_j \leq 1$ by construction. Note that (3) assumes that producers face input prices as given.

2.2. Proposed bootstrap algorithm forcost efficiency

Similarly to input-oriented efficiency scores, [Fare, Grosskopf, Lovell, 1985] cost efficiency scores are linked to $\hat{L}^{\partial}(\mathbf{y})$ and therefore are upwards-biased. Yet, a direct modification of the [Simar, Wilson, 1998; 2007] algorithm to bias correction of cost efficiency score γ, which simply replaces θ by γ at steps 2.2 or step 3.3 (e.g. as proposed in [Borger, Kerstens, Staat, 2008], is inconsistent. Indeed, let us look at a given observation i with coordinates $\mathbf{x}_i$ (point P at Figure 1). By definition of input-oriented efficiency, point $\overline{P}$, which is an intersection of the ray from the origin to P and $\hat{L}^{\partial}(\mathbf{y})$, has coordinates $\hat{\theta}\mathbf{x}_i$. The hyperplane, set by the cost function $\mathbf{w}_i \mathbf{x}_i$ and tangent to $\hat{L}^{\partial}(\mathbf{y})$, intersects the ray from the origin to point P at point P'. Since points P^* and P' are on the same hyperplane, the costs in these points are equal. Therefore, by definition of cost efficiency score, point P' has coordinates $\hat{\gamma}\mathbf{x}_i$. Consequently, point P'', obtained through rescaling inputs by $\hat{\gamma}_i / \hat{\gamma}_{i,b}^*$, belongs to $[P', P]$. However, it may happen that $P'' \notin [\overline{P}, P]$, i.e. $P'' \in [P', P'']$. So the vector of bootstrapped inputs, obtained at step 2.2 of a direct modification of the Simar and Wilson (1998) algorithm, may be outside $\hat{L}(\mathbf{y})$. (The same argument applies to step (3.3) for the case with environmental variables, where $\hat{\theta} = \hat{\theta}(\mathbf{z}_i)$ and $\hat{\gamma} = \hat{\gamma}(\mathbf{z}_i^{\gamma})$. Note that the assumptions about strict convexity of L(y) and free disposability of inputs are importantly exploited in our argument.

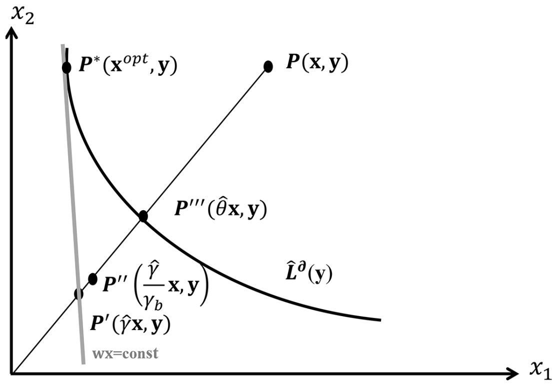

Fig. 1. Bias correction of the cost efficiency, isoquant in the two-input space

To correct for the bias of the [Fare, Grosskopf, Lovell, 1985] cost efficiency we propose the following bootstrap, based on the premises that data generating process for the observed firms satisfy the [Kneip, Simar, Wilson, 2008; Simar, Wilson, 2000b] assumptions 4–6, which we mentioned in the Introduction and which allow a consistent approximation of the unknown distribution of efficiency scores. The bootstrap is homogeneous both in terms of $\hat{f}_\theta(\cdot)$ and $\hat{f}_\gamma(\cdot)$ and constructs pseudo-samples through re-sampling the input-oriented technical efficiency score and rescaling original inputs by the ratio $\hat{\theta}_i / \theta^*_{ib}$. In this way, the bootstrapped inputs are "pushed" inside the $\hat{L}(\mathbf{y})$. Therefore, $\hat{\gamma}^*_{ib}$, which calculated for the bootstrapped inputs at step (4) of our algorithm, allow for consistent bias correction.

1. Estimate naive cost efficiency scores $\hat{\gamma}_1, ..., \hat{\gamma}_J$ for each $i = 1, ..., J$. Assume $(\gamma_1, \ldots, \gamma_J)$ are i.i.d. with pdf $f_\gamma(\cdot)$.
2. Estimate naive input-oriented efficiency scores $\hat{\theta}_1, ..., \hat{\theta}_J$. Assume ($\theta_1, ..., \theta_J$) are i.i.d. with pdf $f_\theta(\cdot)$.
3. Obtain θ^*_{ib} through smoothed bootstrap, and under the assumptions of homogeneous distribution of joint density of θ and joint density of γ in input-output space, assign $\boldsymbol{x}^*_{ib} = \frac{\hat{\theta}_i}{\theta^*_{ib}} \boldsymbol{x}_i, b = 1, \ldots, B$.
4. Calculate $\hat{\gamma}^*_{ib}$ for (x^*_{ib}, y_i).
5. For each i, $\widehat{\text{bias}}\hat{\gamma}_i = \frac{1}{B}\sum_{b=1}^{B}\hat{\gamma}^*_{ib} - \hat{\gamma}_i$

Next, we proceed to the model with environmental variables. We consider a vector zi of environmental variables for the input-oriented model and the vector $\mathbf{z}^\gamma_i$ of environmental variables for the cost minimization model. The corresponding coefficients in the truncated regression for each model are denoted, respectively, as $\boldsymbol{\beta}^\gamma$. Note that $\{\mathbf{z}_i\} \subset \{\mathbf{z}^\gamma_i\}$.

Indeed, as one of the reasons for the bias of cost efficiency scores is the bias of input-oriented scores (owing to the empirical estimate of the frontier), the list of predictors for δ^{γ} includes the list of predictors for δ.

We denote δ_i^{γ} as the reciprocal of the efficiency score in the cost minimization model in order to distinguish it from δ_i , which is the reciprocal of the efficiency score in the input-oriented model. Under the [Simar, Wilson, 2007] assumption about separability of **x** and **z** (i.e. the fact that $L^{\partial}(\mathbf{y})$ does not depend on **z**), we propose the following algorithm for the reciprocal of the [Fare, Grosskopf, Lovell, 1985] cost efficiency score δ_i^{γ} .

1. Estimate reciprocals of naive cost efficiency scores $\hat{\delta}_1^{\gamma},...,\hat{\delta}_J^{\gamma}$, for each i = 1,…,J using system (3).Assume $\delta_i^{\gamma} = \mathbf{z}_i^{\gamma}\hat{\boldsymbol{a}}^{\gamma} + \psi_i \geq 1$, where ψ_i are i.i.d. and independent from $\mathbf{z}_i^{\gamma}, \psi_i \sim N(0,\sigma_{\psi}^2)$ with left truncation at $(1-\mathbf{z}_i^{\gamma}\hat{\boldsymbol{a}}^{\gamma})$.

2. Estimate naive input-oriented distance function scores $\hat{\delta}_1,\dots,\hat{\delta}_J$, for each $i = 1,\dots,J$, using the equivalent of system (2) for reciprocals of θ . Assume, $\delta_i = \mathbf{z}_i\hat{\boldsymbol{a}} + \varepsilon_i \geq 1$, where ε_i are i.i.d. and independent from $\mathbf{z}_i$, $\varepsilon_i \sim N(0,\sigma_{\varepsilon}^2)$ with left truncation at $(1-\mathbf{z}_i\hat{\boldsymbol{a}})$.
3. Use observations for which $\hat{\delta} > 1$ to obtain $\hat{\boldsymbol{a}}$ and $\hat{\sigma}_{\varepsilon}$ in the truncated regression $\hat{\delta}_i = \mathbf{z}_i\hat{\boldsymbol{a}} + \varepsilon_i \geq 1$.
   4. Loop B times to obtain J sets of bootstrap estimates $\{\hat{\delta}_{ib}^{*}\}$, $b = 1,...,B$.
   1. For each $i = 1,...,J$ draw ε_i from $N(0,\hat{\sigma}_{\varepsilon}^2)$ with left truncation at $(1-\mathbf{z}_i\hat{\boldsymbol{a}})$.
2. For each $i = 1,...,J$ compute $(\delta_i^{*} = \mathbf{z}_i\hat{\boldsymbol{a}} + \varepsilon_i)$.
3. Given the semi-parametric dependence of δ on $\mathbf{z}$, assign $\boldsymbol{x}_{ib}^{*} = \frac{\delta_{ib}^{*}}{\hat{\delta}_i}\boldsymbol{x}_i$.
   4. Calculate $\hat{\delta}_{ib}^{\gamma *}$ for $(\mathbf{x}_{ib}^{*},\mathbf{y}_i)$.
5. Owing to semi-parametric dependence of δ on **z**, we can compute $\text{bias}\hat{\hat{\delta}}_i^{\gamma} = \frac{1}{B}\sum_{b=1}^{B}\hat{\delta}_{ib}^{\gamma *} - \hat{\delta}_i^{\gamma}$ and bias-corrected score $\hat{\hat{\delta}}_i^{\gamma} = \hat{\delta}_i^{\gamma} - \text{bias}\hat{\hat{\delta}}_i^{\gamma}$

2.3. Naive cost efficiency score with input prices under producer control

Tone (2002) concentrates on input costs, assuming that producers may choose prices for their inputs. Let $\overline{\mathbf{x}}_j = (w_{1j}x_{1j},...,w_{Nj}x_{Nj})^T$, $\overline{\mathbf{X}} = (\overline{\mathbf{x}}_1,...,\overline{\mathbf{x}}_J)^T$, where $\mathbf{w}_j$ is a vector of prices for each input xj. The Tone (2002) “new” cost efficiency for DMU j is defined as

$$\overline{\gamma}_j = \mathbf{e}\overline{\mathbf{x}}_j^{opt} / \mathbf{e}\overline{\mathbf{x}}_j$$

with $\overline{\mathbf{x}}_j^{opt}$ a solution to (constant returns to scale formulation):

$$\min_{\overline{\mathbf{x}}_j,\lambda} \mathbf{e}\overline{\mathbf{x}}_j$$

$$s.t. -y_{mj} + \sum_{i=1}^{J}\lambda_i y_{mi} \geq 0,\ \ m = 1,\dots,M,\ \ (5)$$

$$\bar{x}_{nj} - \sum_{i=1}^{J} \lambda_i \bar{x}_{ni} \geq 0, \;\; n = 1, \ldots, N,$$

$$\lambda_i \geq 0, \;\; i = 1, \ldots, J.$$

Here **e** is a unit vector, and by construction in (4) and (5), $0 \leq \bar{\gamma}_j \leq 1$.

2.4. Proposed bootstrap algorithm for the new cost efficiency

Denote T_n technology in [Tone, 2002] of "new" technical (and cost) efficiency estimates.

$$T_n = \left\{ (\bar{\boldsymbol{x}}, \boldsymbol{y}) : \bar{\boldsymbol{x}} \geq \bar{\boldsymbol{X}} \boldsymbol{\lambda}, \boldsymbol{y} \leq \boldsymbol{Y\lambda}, \boldsymbol{\lambda} \geq 0 \right\} \tag{6}$$

Define the "new" input set $L_n(\mathbf{y}) = \{(\bar{\mathbf{x}}) : (\bar{\mathbf{x}}, \mathbf{y}) \in T_n\}$. As is demonstrated in [Tone, 2002] (theorem 4), the set of constraints on each $\bar{x}_{nj}$ in (5) is equivalent to this aggregate constraint:

$$\boldsymbol{e}\,\bar{\boldsymbol{x}} - \boldsymbol{e}\bar{\boldsymbol{X}}\boldsymbol{\lambda} \geq 0. \tag{7}$$

Consequently, for a given level of $\mathbf{y}$, the $\hat{L}_n^{\partial}(\mathbf{y})$ is a hyperplane, parallel to the hyperplane set by a given level of the objective function $\mathbf{e}\bar{\mathbf{x}}_j$. Therefore, the tangency of the objective function and $\hat{L}_n^{\partial}(\mathbf{y})$ implies that the two hyperplanes are coincident (Figure 2). Accordingly, the ray from origin to the point $P \in \hat{L}_n(\mathbf{y})$ intersects $\hat{L}_n^{\partial}(\mathbf{y})$ and the hyperplane, set by the objective function, at the same point. So $P' = \bar{P}$. In other words, as is noted in theorem 6 in Tone (2002), the "new" cost efficiency point is also "new" technically efficient.[4] So a consistent bias correction of the Tone (2002) "new" cost efficiency score may be conducted through a direct application of the Simar and Wilson (1998) and

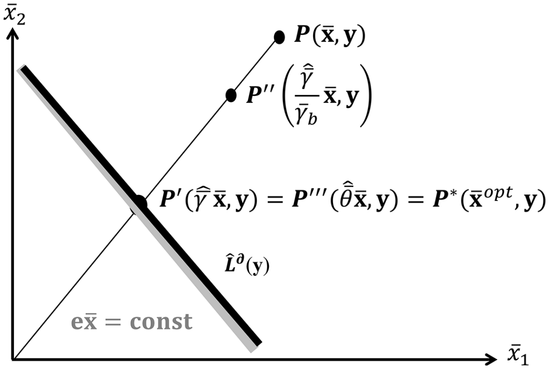

Fig. 2. Bias correction of the newcost efficiency, isoquant in two-input space

[4] Therefore, papers that estimate input-oriented efficiency scores using input costs as inputs and interpret the scores as cost efficiency (e. g. [Medin et al., 2011; Linna et al., 2010; Barros, Dieke, 2008] in fact, measure the [Tone, 2002]) "new" cost efficiency.

the Simar and Wilson (2007) algorithm, so that the following rescaling is implemented at step (3.3): $x^*_{ib} = \frac{\overline{\gamma}_i}{\overline{\gamma}^*_{ib}} \overline{x}_i$. Indeed, as $L^{\partial}_n(\mathbf{y})$ is set by the aggregate constraint (7), $P'' \in [P', P]$ is equivalent to $P'' \in [\overline{P}, P]$. Therefore, rescaling guarantees that each component of $\overline{\mathbf{x}}_b$ is larger than the corresponding component of the original vector $\overline{\mathbf{x}}$, and vector $\overline{\mathbf{x}}_b$ lies in the necessary subspace relative to $L^{\partial}_n(\mathbf{y})$ [Besstremyannaya, 2013].

3. Simulations

3.1. Microeconomic framework

The Cobb–Douglas production function, commonly used in the non-parametric efficiency analysis in the banking industry [Kneip, Simar, Wilson; 2008, 2011; Fethi, Pasiouras, 2010; Thanassoulis, Portela, Despic 2008; Badin, Simar,2003; Simar, Wilson, 2002; 2000b; Kittelsen,1999; Banker, Gadh, Gorr, 1993] is taken in the form of [Kumbhakar, 2011; Resti, 2000]

$$y_m = A_m \prod_{n=1}^{N} x_{nm}^{\alpha_{nm}}, \tag{8}$$

where x_{nm} is the quantity of n-th input, used to produce m-th output ($x_n = \sum_{m=1}^{M} x_{nm}$), A_m and α_{nm} are the parameters. Outputs y_m and input prices w_n are assumed to come from multivariate lognormal distribution, where vector of means and variance-covariance matrix are taken from our real banking data (in particular, are based on the asset approach in defining the input and output pairs)[5]. The minimal dimension of the input vector, required for differentiating between technical and cost efficiency, is two. Yet, banking is commonly considered to be a multi-output industry, and so we exploit two-output and three-input models:

$$\ln(\boldsymbol{y}) \sim \left(\begin{pmatrix} 7.3498 \\ 6.2898 \end{pmatrix}, \begin{pmatrix} 1.2686 & 1.4686 \\ 1.4680 & 1.8260 \end{pmatrix} \right) \text{and} \ln(\boldsymbol{w}) \sim \left(\begin{pmatrix} -4.9157 \\ -2.0093 \\ -5.5727 \end{pmatrix}, \begin{pmatrix} 0.0309 & 0.0368 & 0.0231 \\ 0.0368 & 0.2079 & 0.0702 \\ 0.0231 & 0.0702 & 0.1193 \end{pmatrix} \right).$$

In the absence of environmental variables, we employ the [Resti, 2000] approach of introducing cost inefficiencies to $(N-1)$ inputs and analytically computing the value of the N-th input, so that the firm remained on the same isoquant (with unchanged level of input-oriented efficiency): $x_{nm} = x^*_{nm}\eta_{nm}$, where $n = 1, \ldots, N-1$; $\eta_{nm} > 0$. Then, cost efficiency γ is calculated as follows (Appendix, eq. A. 13):

[5] To check for robustness, we conducted a second set of simulations with data from the intermediation approach, and we found similar results.

$$\gamma = \frac{wx^{opt}}{wx} = \frac{\sum_{n=1}^{N}\sum_{m=1}^{M}(y_m^{*1/\rho_m} / A_m)\theta\alpha_{nm}T_m}{\sum_{n=1}^{N-1}\sum_{m=1}^{M}(y_m^{*1/\rho_m} / A_m)\alpha_{nm}T_m\eta_{nm} + \sum_{m=1}^{M}(y_m^{*1/\rho_m} / A_m)\alpha_{Nm}T_m\prod_{n=1}^{N-1}\eta_{nm}^{-\alpha_{nm}/\alpha_{Nm}}}, \quad (9)$$

where $\rho_m = \sum_{n=1}^{N}\alpha_{nm}$ and $T_m = \prod_{n=1}^{N}\left(\frac{w_n}{\alpha_{nm}}\right)^{\alpha_{nm}/\rho_m}$.

Our estimations with Japanese data and results from the empirical literature [Liu, Ondrich, Ruggiero, 2012; Wang, 2003; Banker, Gadh, Gorr, 1993; Giokas, 1991] show that input elasticities do not vary appreciably for banking outputs employed in this paper. Hence, we regard $\alpha_{nm} \equiv \alpha_n$ (and hence $\rho_m \equiv \rho$) and $\eta_{nm} \equiv \eta_n$. Accordingly, we obtain (eq.A.15):

$$\gamma = \frac{\rho\theta}{\sum_{n=1}^{N-1}\alpha_n\eta_n + \alpha_N\prod_{n=1}^{N-1}\eta_n^{-\alpha_n/\alpha_N}}. \quad (10)$$

We use constant returns to scale with $(\alpha_1, \alpha_2, \alpha_3) = (0.05, 0.05, 0.9)$. Inefficiencies are added, so that $\mathbf{y} = \mathbf{y}^*\theta^{\rho}$, $0 < \theta^{\rho} \le 1$ (see [Kneip, Simar, Wilson 2011; Badin, Simar 2003; Simar, Wilson, 2002; 2000b; Resti, 2000; Kittelsen, 1999]). Then, owing to homothetic property of cost function, the input-oriented efficiency is θ. Cost inefficiencies are added to x_2 and x_3 and analytically computed for x_1. Input-oriented efficiency is $\theta = 1/(1+\zeta)$, where ζ is drawn from $Exp(2)$ and $E(\zeta) = 0.5$. Note that $1 + Exp(2)$ has high probability of obtaining a point in the neighborhood of unity. Consequently, the *DGP* with exponential distribution allows easier estimation of the frontier if compared to *DGPs* with fewer points in the proximity of unity6. In the presence of environmental variables, we introduce inefficiencies as $\mathbf{y} = \mathbf{y}^*\delta^{-\rho}$, $0 < \delta^{-\rho} \le 1$, where δ can be expressed as $\boldsymbol{z\beta} + \varepsilon$. We assume a simplified case when the lists of environmental variables, influencing input-oriented efficiency and cost efficiency, coincide. $\delta \sim N(\mu_z, \sigma_z^2)$ with left truncation at unity. Following Simar and Wilson (2007), we set $r = 2$, $\beta_1 = \beta_2 = 0.5$, $z_1 = 1$, $z_2 \sim N(2,4)$, $\varepsilon \sim N(0,1)$, with left-truncation at $(1 - \boldsymbol{z})$, $\boldsymbol{\delta} = \boldsymbol{z\hat{a}} + \boldsymbol{\varepsilon}$. Then eq. (10) modifies to

$$\delta^{\gamma}(\mathbf{z}) = \frac{\delta}{\rho}\left(\sum_{n=1}^{N-1}\alpha_n\eta_n + \alpha_N\prod_{n=1}^{N-1}\eta_n^{-(\alpha_n/\alpha_N)}\right).$$

As regards cost efficiency, $\eta_n = e^{\nu_n}$, where $\nu_n \sim N(0, \sigma_\nu^2)$. In this case the realized value of η_n may be smaller or larger than unity, and it allows to move $\mathbf{x}^*$ in different directions along the isoquant. To model different size of cost inefficiencies, we take $\sigma_\nu = \{0.05, 0.1\}$. Following [Simar, Wilson, 2011b], we use 1000 trials with B=2000 iterations on each trial. Our samples $J = \{50, 100, 200, 300, 400, 600, 800, 1000\}$ and confidence levels $(1-\alpha)$ are 0.9, 0.95 and 0.99.

[6] If data-generating process results in a small number of points in the proximity of unity, the consistent estimation of the frontier would require increasing sample size appreciably.

A fixed point to measure cost efficiency on each trial is constructed as follows. We take a vector in the middle of the output and price data and assign it input-oriented efficiency $E\theta$. The coordinates of a point on the frontier are $\left(\mathbf{x}^*\left([E\theta]^\rho \boldsymbol{i}_y, \boldsymbol{i}_w\right), [E\theta]^\rho \boldsymbol{i}_y\right)$, where $\mathbf{x}^*(\cdot,\cdot)$ is a an optimal demand function from eq.(A.2). Then, we introduce inefficiencies $E\eta$ to $(N-1)$ input coordinates of the point, and analytically compute the values of N-th input coordinate according to eq.(A.5).

3.2. Results

Owing to potential problems of ignoring the zero bound in implementing the [Silverman, 1986] reflection method with the input-oriented efficiency scores θ [Simar, Wilson, 2000a], the estimations are conducted in terms of the reciprocals $\delta = 1/\theta$. Accordingly: first, each point $\hat{\delta}_i \geq 1$ is reflected by its symmetric image $2-\hat{\delta}_i \leq 1$; second, kernel density is estimated from the set of $2J$ points [Simar , Wilson, 2008]. Our estimates demonstrate that the absolute difference between the true and bias-corrected values of cost efficiency both in absence and presence of environmental variables is close to 0.04 with the smallest sample size ($J=50$) and becomes less than 0.01 with $\mathrm{J} > 600$ (Tables 1–2). As the values of efficiency scores belong to the [0, 1] segment, it may be concluded that our bias correction on average leads to the estimates with negligible difference from the true value under moderate sample sizes and reasonable numbers of bootstrap iterations.

Table 1. **Absolute difference between the true and estimated cost efficiency for homogeneous smooth bootstrap in the absence of environmental variables, with sample adjusted cross-validation bandwidth**

J	σ_v	$\alpha=1$		$\alpha=0.05$		$\alpha=1.0$	
50	0.05	0.036	[0.012]	0.038	[0.012]	0.037	[0.012]
100	0.05	0.026	[0.008]	0.026	[0.008]	0.026	[0.008]
200	0.05	0.018	[0.005]	0.018	[0.005]	0.018	[0.005]
300	0.05	0.014	[0.004]	0.014	[0.004]	0.014	[0.004]
400	0.05	0.012	[0.003]	0.012	[0.003]	0.012	[0.003]
600	0.05	0.01	[0.002]	0.01	[0.002]	0.01	[0.002]
800	0.05	0.008	[0.002]	0.009	[0.002]	0.008	[0.002]
1000	0.05	0.007	[0.002]	0.007	[0.002]	0.008	[0.002]
50	0.1	0.037	[0.012]	0.037	[0.012]	0.037	[0.012]
100	0.1	0.025	[0.008]	0.026	[0.008]	0.025	[0.008]
200	0.1	0.018	[0.005]	0.018	[0.005]	0.018	[0.005]
300	0.1	0.014	[0.004]	0.014	[0.004]	0.014	[0.004]

J	σ_v	$\alpha = 1$		$\alpha = 0.05$		$\alpha = 1.0$	
400	0.1	0.012	[0.003]	0.012	[0.003]	0.012	[0.003]
600	0.1	0.01	[0.002]	0.01	[0.002]	0.01	[0.002]
800	0.1	0.008	[0.002]	0.008	[0.002]	0.008	[0.002]
1.000	0.1	0.007	[0.002]	0.008	[0.002]	0.007	[0.002]

Note: standard deviation in brackets.

Table 2. **Absolute difference between the true and estimated cost efficiency for semi-parametric bootstrap in the presence of environmental variables**

J	σ_v	$\alpha = 1$		$\alpha = 0.05$		$\alpha = 1.0$	
50	0.05	0.043	[0.017]	0.043	[0.017]	0.042	[0.016]
100	0.05	0.029	[0.010]	0.03	[0.010]	0.029	[0.010]
200	0.05	0.02	[0.006]	0.02	[0.005]	0.02	[0.005]
300	0.05	0.016	[0.004]	0.016	[0.004]	0.016	[0.004]
400	0.05	0.014	[0.003]	0.014	[0.003]	0.014	[0.003]
600	0.05	0.011	[0.002]	0.011	[0.002]	0.011	[0.002]
800	0.05	0.01	[0.002]	0.01	[0.002]	0.01	[0.002]
1.000	0.05	0.009	[0.002]	0.009	[0.002]	0.009	[0.002]
50	0.1	0.043	[0.018]	0.042	[0.017]	0.042	[0.016]
100	0.1	0.029	[0.010]	0.03	[0.010]	0.03	[0.010]
200	0.1	0.021	[0.006]	0.021	[0.006]	0.021	[0.006]
300	0.1	0.017	[0.004]	0.017	[0.004]	0.017	[0.004]
400	0.1	0.014	[0.004]	0.014	[0.003]	0.014	[0.003]
600	0.1	0.012	[0.003]	0.012	[0.003]	0.012	[0.003]
800	0.1	0.01	[0.002]	0.01	[0.002]	0.01	[0.002]
1.000	0.1	0.009	[0.002]	0.009	[0.002]	0.009	[0.002]

Notes: standard deviation in brackets; $dim(\boldsymbol{x}) = dim(\boldsymbol{y}) = 2$.

4. Efficiency estimates for Japanese banks

4.1. Data

We use data from the Japanese Bankers Association, which provides financial variables for all Japanese banks from their consolidated financial statements and statements of cash flow, along with the number of employees, bank branches, and bank charters from interim financial statements. Regional (prefectural) variables come from the Bank of Japan (deposits, vault cash, loans, and bills discounted), Economic and Social Research Institute, Cabinet Office (gross domestic product and gross domestic product deflator), Ministry of Land, Infrastructure and Transport, and Japan Statistical Yearbook (price of commercial land site).

Following common approaches to cost efficiency analyses in banking, we exploit a three input-two output model, where outputs are either performing loans and total securities (asset approach, e. g. [Hori, Yoshida, 1996; Fukuyama, Weber, 2002; Barros, Managi, Matousek, 2012) or revenue from loans and revenue from other business activities (intermediation approach, e. g. [Kasuya, 1986; Fukuyama,1993; 1995; Takahashi, 2000; Fukuyama, Weber, 2010; Thanassoulis, Portela, Despic, 2008; Tortosa-Austina, 2002]). In each model the inputs are labor (total employees), capital (premises, real estate, and intangibles) and funds from customers (we follow [Kasuya, 1986; 1989; Fukuyama, 1993; 1995; Hori, Yoshida, 1996; McKillop, Glass, Morikawa, 1996; Glass, McKillop, Morikawa, 1998; Fukuyama, Weber, 2002; Miyakoshi, Tsukuda, 2004; Fukuyama, Weber, 2008; Barros, Managi, Matousek, 2012]). The proxies for input prices are, respectively, personnel expenditure/total employees, capital expenditure/capital, and fund-raising expenditure/funds from customers (e. g. as in [Kasuya, 1986; 1989; McKillop, Glass, Morikawa, 1996; Fukuyama, Weber, 2002]). The choice of inputs, outputs, and prices follows the methodology of efficiency analysis in Japanese banking[7]. Bank-level environmental variables include bank size and bank product diversity (see [Aly et al., 1990; Simar, Wilson, 2007]), ratio of loan loss provisions to total loans (following [Altunbas et al., 2000; Drake, Hall, 2003; Drake, Hall, Simper, 2009])[8]. Prefecture-level environmental variables are share of monetary aggregate in gross regional product, real rate of growth of gross domestic product, and commercial land price [Liu, Tone, 2008]. We include dichotomous variables by bank charter (city bank, regional bank, regional second tier bank, trust bank, long-term credit bank). Bank holdings and financial groups are excluded from the analysis, as they may have zero reported capital (Table 3).

Our sample uses data for the fiscal year which runs from April 2013 to March 2014. The sample represents the entire banking industry in Japan, yet its size is only 106. However, the results of our simulations demonstrate relatively small differences between the true and the estimated cost efficiency, even for such small samples.

Table 3. **Descriptive statistics for the sample of 106 Japanese banks in fiscal year 2013**

Variable	Definition	Mean	Std. Dev.	Min	Max
Inputs					
x_1	labor = total employees (including board)	2714	4481	312	31461
x_2	capital = premises and real estate + intangibles	26	123	0.044	1125

[7] Note that the intermediation approach prevails in international literature [Fethi, Pasiouras, 2010], yet the asset approach is more widespread in analyses of Japanese banking. See review of the literature on measuring the efficiency of Japanese banks in [Besstremyannaya, 2017].

[8] Using the non-performing loans in an alternative approach does not change the results of the estimates appreciably, since loan loss provisions and non-performing loans are highly correlated.

Variable	Definition	Mean	Std. Dev.	Min	Max
x_3	funds from customers = total deposits + negotiable certificates of deposits + call money + bills sold + borrowed money + foreign exchange deposits + other deposits	7366	21589	219	157288
Outputs					
Asset approach					
y_1	performing loans = total loans — nonperforming loans	4614	12566	158	89543
y_2	securities and other interest bearing assets	2575	8434	1.4	66543
Intermediation approach					
y_3	revenue from loans = interest on loans and discounts + interest on bills bought	67	181	3.6	1326
y_4	revenue from other business activity = total operating income — other operating income — interest and dividends on securities — y_3	63	214	0.706	1428
Input prices					
w_1	labor price = (general and administrative expenses-depreciation)/total employees	0.017	0.007	0.009	0.065
w_2	capital price = (expenditure on premises and fixed assets)/x_2	1.126	1.399	0.136	10.442
w_3	price of funds = fund raising expenditure/x_3	0.001	0.001	0	0.005
Bank variables					
z_1	= ln(branches)	4.56	0.62	3.05	6.72
z_2	Herfindahl index of product diversity	0.72	0.23	0.26	1.44
z_3	nonperforming loan ratio = nonperforming loans /total loans	0.03	0.01	0.004	0.07
z_4	= 1 if city bank	0.05	0.21	0	1
z_5	= 1 if regional bank	0.56	0.5	0	1
z_6	= 1 if regional tier 2 (former Sogo) bank	0.35	0.48	0	1
z_7	= 1 if trust bank	0.03	0.17	0	1
z_8	= 1 if longterm credit bank	0.02	0.14	0	1

Variable	Definition	Mean	Std. Dev.	Min	Max
Prefectural variables					
z_9	rate of growth of gross regional product (in 2010 real terms)	1	0.05	0.93	1.18
z_10	share of monetary aggregate (M2 + negotiable certificates of deposit) in regional product	0.7	0.32	0.4	1.52
z_11	share of loans in gross regional product	0.75	0.46	0.36	1.95
z_12	rate of growth of price of commercial land (in 2010 real terms)	0.58	0.14	0.41	0.92

Note: financial variables are in billion yen. The sample size is 106.

4.2. Results

Estimations are conducted under variable returns to scale with B=2000. We exploit least squares cross-validation bandwidth and use $z1-z3$, $z9-z12$ and a dummy for city banks ($z4$) in the model with the environmental variables. (The remaining dichotomous variables for other bank charters are omitted, owing to multicollinearity).

Table 4. **Cost efficiency scores for Japanese banks**

Score		Asset approach	Intermediation approach
$\hat{\gamma}$	mean	0.7189	0.6679
	st.dev.	0.1406	0.166
	range	[0.4496, 1]	[0.3781, 1]
$\hat{\hat{\gamma}}$	mean	0.6622	0.5906
	st.dev.	0.13	0.137
	range	[0.4091, 0.9503]	[0.3338, 0.9011]
$1/\hat{\delta}^{\gamma}$	mean	0.7276	0.6649
	st.dev.	0.1301	0.1544
	range	[0.5130, 1]	[0.4220, 1]
$1/\hat{\hat{\delta}}^{\gamma}$	mean	0.6737	0.615
	st.dev.	0.1187	0.1205
	range	[0.4832, 0.9732]	[0.3739, 0.8928]

Table 4 shows the estimates of "naïve" score $\hat{\gamma}$ ($1/\hat{\delta}^{\gamma}$) and bias-corrected score $\hat{\hat{\gamma}}$ ($1/\hat{\hat{\delta}}^{\gamma}$) for the models, corresponding to the asset approach and the intermediation approach. In each model, the mean bias-corrected score is lower than the mean "naïve" score, while the standard deviation of the "naïve" and bias-corrected scores are close. The bias-corrected score is "to the left" (if compared to the range of the "naïve" score), and there are no exact unity values of bias-corrected cost efficiency. The mean value of cost efficiency is higher in the model, with the asset approach both in presence and in absence of environmental variables. Accounting for environmental variables leads to higher cost efficiency scores, if compared to corresponding models without environmental variables.

Conclusion

The paper shows that a direct modification of the [Simar, Wilson, 1998; 2007] methodology is inconsistent for correcting the bias of the [Fare, Grosskopf, Lovell, 1985] cost efficiency scores and proposes an alternative bootstrap algorithm for estimation. To approximate the bias of the "naïve" cost efficiency score, the proposed algorithm re-samples "naïve" input-oriented efficiency scores, rescales original inputs to bring them to the frontier, and then re-estimates cost efficiency scores for the rescaled inputs.

The results of simulation analyses for a multi-input, multi-output Cobb–Douglas production function with correlated outputs, and correlated technical and cost efficiency, show consistency of the proposed algorithm. An application of the algorithm to real data of 106 Japanese banks for fiscal year 2013 allows quantifying the bias of the naïve cost efficiency scores.

References

Altunbas Y., Liu M-H., Molyneux P., Seth R. (2000) Efficiency and Risk in Japanese Banking. *Journal of Banking and Finance*, vol. 24, iss. 10, pp. 1605–28.

Aly H. Y., Grabowski R., Pasurka C., Rangan N. (1990) Technical, Scale, and Allocative Efficiencies in U. S. Banking: An Empirical Investigation. *Review of Economics and Statistics*, no. 72, pp. 211–17.

Badin L., Simar L. (2003) Confidence Intervals for DEA-Type Efficiency Scores: How to Avoid the Computational Burden of the Bootstrap. Technical report 0322. Available at: http://sites.uclouvain.be/IAP-Stat-Phase-V-VI/PhaseV/publications_2003/TR/TR0322.pdf: IAP Statistics Network (accessed: 06.03. 2019).

Banker R., Gadh V. M., Gorr W. L. (1993) A Monte Carlo Comparison of Two Production Frontier Estimation Methods: Corrected Ordinary Least Squares and Data Envelopment Analysis. *European Journal of Operational Research*, no. 67, pp. 332–43.

Barros C. P., Dieke P. U. C. (2008) Measuring the Economic Efficiency of Airports: A Simar-Wilson Methodology Analysis. *Transportation Research.* Part E, no. 44, pp. 1039–51.

Barros C. P., Managi S., Matousek R. (2012) The Technical Efficiency of the Japanese Banks: Non-Radial Directional Performance Measurement with Undesirable Output. *Omega*, no. 40, pp. 1–8.

Besstremyannaya G. (2013) The Impact of Japanese Hospital Financing Reform on Hospital Efficiency. *Japanese Economic Review*, vol. 64, iss. 3, pp. 337–362.

Besstremyannaya G. (2017) Heterogeneous Effect of the Global Financial Crisis and the Great East Japan Earthquake on Costs of Japanese Banks. *Journal of Empirical Finance*, no. 42, pp. 66–89.

Borger B. de, Kerstens K., Staat M. (2008) Transit Costs and Cost Efficiency: Bootstrapping Non-Parametric Frontiers. *Research in Transportation Economics*, no. 23, pp. 53–64.

Charnes A., Cooper W., Rhodes E. (1978) Measuring the Efficiency of Decision Making Units. *European Journal of Operational Research*, no. 2, pp. 429–44.

Coelli T. J., Rao D. S. P., Battese G. E. (1994) *An Introduction to Efficiency and Productivity Analysis*. Kluwer Academic Publishers.

Drake L., Hall M. J. B. (2003) Efficiency in Japanese Banking: An Empirical Analysis. *Journal of Banking and Finance*, no. 27, pp. 891–917.
Drake L., Hall M. J. B., Simper R. (2009) Bank Modelling Methodologies: A Comparative Non-Parametric Analysis of Efficiency in the Japanese Banking Sector. *Journal of International Financial Markets, Institutions and Money*, no. 19, pp. 1–15.
Fare R., Grosskopf S., Lovell C. A. (1985) *The Measurement of Efficiency in Production*. Boston: Kluwer Academic Publishers.
Farrell M. J. (1957) The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society*, Series A (General), no. 3, pp. 253–81.
Fethi M. D., Pasiouras F. (2010) Assessing Bank Efficiency and Performance with Opeartional Research and Artificial Intelligence Techniques: A Survey. *European Journal of Operational Research*, no. 204, pp. 189–98.
Fukuyama H. (1993) Technical and Scale Efficiency of Japanese Commercial Banks: A Non-Parametric Approach. *Applied Economics*, no. 25, pp. 1101–1112.
Fukuyama H. (1995) Measuring Efficiency and Pruductivity Growth in Japanese Banking: A Nonparametric Frontier Approach. *Applied Financial Economics*, no. 5, pp. 95–107.
Fukuyama H., Weber W. L. (2002) Estimating Output Allocative Efficiency and Productivity Change: Application to Japanese Banks. *European Journal of Operation Research*, no. 137, pp. 177–90.
Fukuyama H., Weber W. L. (2008) Japanese Banking Inefficiency and Shadow Pricing. *Mathematical and Computer Modelling*, no. 48, pp. 1854–67.
Fukuyama H., Weber W. L. (2010) A Slacks-Based Inefficiency Measure for a Two-Stage System with Bad Outputs. *Omega*, no. 38, pp. 398–409.
Giokas D. (1991) Bank Branch Operating Efficiency: A Comparative Application of DEA and the Loglinear Model. *Omega*, vol. 19, iss. 6, pp. 549–57.
Glass J. C., McKillop D. G., Morikawa Y. (1998) Intermedication and Value-Added Models for Estimating Cost Economies in Large Japanese Banks 1977–93. *Applied Financial Economics*, no. 8, pp. 653–61.
Hori K., Yoshida A. (1996) Nihon-No Ginkougyou-No Hiyou Kouritsusei [Measuring Cost Efficiency in Japanese Banking]. *Japanese Journal of Financial Economics*, no. 1, pp. 87–110.
Kasuya M. (1986) Economies of Scope: Theory and Application to Banking. *BOJ Monetary and Economic Studies*, no. 4, pp. 59–104.
Kasuya M. (1989) Ginkou-No Kosutokouzou-No Jisshobunseki. Kouritsusei, Gijutsusinpo, Yosoukankawase-Ni Kansuru Gyoutaibetsu Jisshobunseki [Empirical Analysis of Cost Structure in Banking. Empirical Analysis of Efficiency, Technological Progress and Input Substitution by Industry]. *Kinyuu Kenkyuu*, no. 8, pp. 79–118.
Kittelsen S. A. C. (1999) Monte Carlo Simulations of DEA Efficiency Measures and Hypothesis Tests. *Memorandum 09*. Department of Economics, University of Oslo.
Kneip A., Simar L., Wilson P. W. (2008) Asymptotics and Consistent Bootstraps for DEA Estimators in Nonparametric Frontier Models. *Econometric Theory*, no. 24, pp. 1663–97.
Kneip A., Simar L., Wilson P. W. (2011) A Computationally Efficient, Consistent Bootstrap for Inference with Non-Parametric DEA Estimators. *Computational Economics*, no. 38, pp. 483–515.
Kumbhakar S. C. (2011) Estimation of Multiple Output Production Functions. 2012 spring seminar series April 17. http://www.econ.gatech.edu/files/seminars/kumbhakar11.pdf: Georgia Tech (accessed: 06.03. 2019).
Linna M., Hakkinen U., Peltola M., Magnussen J., Anthun K. S., Kittelsen S., Roed A., Olsen K., Medin E., Rehnberg C. (2010) Measuring Cost Efficiency in the Nordic Hospitals — a Cross-Sectional Comparison of Public Hospitals in 2002. *Health Care Management Science*, no. 13, pp. 346–57.
Liu J., Tone K. (2008) A Multistage Method to Measure Efficiency and Its Application to Japanese Banking Industry. *Socio-Economic Planning Sciences*, no. 42, pp. 75–91.
Liu L., Ondrich J., Ruggiero J. (2012) Estimating Multiple-Input–multiple-Output Production Functions with an Analysis of Credit Unions. *Applied Economics*, no. 44, pp. 1583–1589.
McKillop D. G., Glass J. C., Morikawa Y. (1996) The Composite Cost Function and Efficiency in Giant Japanese Banks. *Journal of Banking and Finance*, no. 20, pp. 1651–1671.
Medin E., Anthum K. S., Hakkinen U., Kittelsen S. A. C., Linna M., Magnussen J., Olsen K., Rehnberg C. (2011) Cost Efficiency of University Hospitals in the Nordic Countries: A Cross-Country Analysis. *European Journal of Health Economics*, vol. 12, iss. 6, pp. 509–519.
Miyakoshi T., Tsukuda Y. (2004) Regional Disparities in Japanese Banking Performance. *RURDS*, vol. 16, iss. 1, pp. 74–89.
Resti A. (2000) Efficiency Measurement for Multi-Product Industries: A Comparison of Classic and Recent Techniques Based on Simulated Data. *European Journal of Operational Research*, no. 121, pp. 559–78.

Shephard R. W. (1981) Cost and Production Functions. *Lecture Notes in Economics and Mathematical Systems*. Ed. by M. Beckmann, H. P. Kunzi. Berlin, Heidelberg, New York, Springer-Verlag.

Silverman B. W. (1986) Density Estimation for Statistics and Data Analysis. New York: Chapman; Hall.

Simar L., Wilson P. (1998) Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models. *Management Science*, no. 44, pp. 49–61.

Simar L., Wilson P. (2000a) A General Methodology for Bootstrapping in Non-Parametric Frontier Models. *Journal of Applied Statistics*, vol. 27, iss. 6, pp. 779–802.

Simar L., Wilson P. (2000b) *Performance of the Bootstrap for DEA Estimators and Iterating the Principle*. STAT Discussion Papers 0002.

Simar L., Wilson P. (2002) Non-Parametric Tests of Returns to Scale. *European Journal of Operational Research*, vol. 139, iss. 1, pp. 115–32.

Simar L., Wilson P. (2007) Estimation and Inference in Two-Stage, Semi-Parametric Models of Production Processes. *Journal of Econometrics*, no. 136, pp. 31–64.

Simar L., Wilson P. (2008) Statistical Inference in Nonparametric Frontier Models: Recent Developments and Perspectives. *The Measurement of Productive Efficiency and Productivity Growth*. Eds H. O. Fried, C. A. K. Lovell, S. S. Schmidt, Oxford University Press, pp. 421–521.

Simar L., Wilson P. (2011a) Inference by the M Out of N Bootstrap in Nonparametric Frontier Models. *Journal of Productivity Analysis*, no. 36, pp. 33–53.

Simar L., Wilson P. (2011b) Two-Stage DEA: Caveat Emptor. *Journal of Productivity Analysis*, no. 36, pp. 205–18.

Simm J., Besstremyannaya G. (2016) Robust Data Envelopment Analysis (DEA) for R. Available at: https://Cran.r-Project.org/Web/Packages/RDEA/Index.html (accessed: 06.03.2019).

Takahashi T. (2000) DEA-Wo Motiiteno Hatansaikenshori-Mo Kamishita Ginkou-No Kouritsusei-No Keisoku [Measurement of Bank Efficiency with DEA and the Analysis of Bad Loans]. *Operations Research*, no. 479, pp. 598–602.

Thanassoulis E., Portela M. C. S., Despic O. (2008) Banking. *The Measurement of Productive Efficiency and Productivity Growth*. Eds H. O. Fried, C. A. K. Lovell, S. S. Schmidt. Oxford University Press.

Tone K. (2002) A Strange Case of the Cost and Allocative Efficiencies in DEA. *Journal of the Operational Research Society*, vol. 53, iss. 11, pp. 1225–31.

Tortosa-Austina E. (2002) Bank Cost Efficiency and Output Specification. *Journal of Productivity Analysis*, no. 18, pp. 199–222.

Wang J. C. (2003) Productivity and Economies of Scale in the Production of Bank Service Value Added. Working Paper September. Available at: http://www.bostonfed.org/economic/wp/wp2003/wp037.pdf: Research Department, Federal Reserve Bank of Boston (accessed: 06.03. 2019).

Received: September 26, 2018
Accepted: December 17, 2018

Author's information:

Galina E. Besstremyannaya — PhD in Economics; gbesstre@nes.ru
Jaak Simm — PhD in Computer Science; jaak.simm@kuleuven.be

Оценка ценовой эффективности в непараметрических моделях производственной границы

Г. Е. Бесстремянная[1], *Я. Симм*[2]

[1] Российская экономическая школа и Институт Дальнего Востока РАН,
Российская Федерация, 121353, Москва, Сколковское шоссе, 45
[2] Stadius Centre for Dynamical Systems,
B–3001, Хеверли, Бельгия

Для цитирования: Besstremyannaya G., Simm J. (2019) Estimation of Cost Efficiency in Non-parametric Frontier Models. *Вестник Санкт-Петербургского университета. Экономика*. Т. 35. Вып. 1. С. 3–25. https://doi.org/10.21638/spbu05.2019.101 (In English)

В работе предлагается методология для несмещенной оценки ценовой эффективности в непараметрических моделях производственный границы (data envelopment analysis).

Рассматривается традиционное определение ценовой эффективности в моделях производственной границы, а также ценовая эффективность, введенная применительно к случаю, когда цены на факторы различаются для каждого производителя. Для ценовой эффективности по Фаре (Fare), Гросскопфу (Grosskopf) и Ловеллу (Lovell) наш алгоритм бутстрапа осуществляет выборки из «наивных» оценок в рамках модели, ориентированной на факторы производства. Затем происходит взвешивание исходных факторов, так что новые факторы оказываются на границе, и делается оценка ценовой эффективности для взвешенных факторов. Для случая ценовой эффективности по Тонэ (Tone) показана применимость бутстрапа из работ Симара (Simar) и Уилсона (Wilson). В статье рассматриваются модели с факторами производства и выпусками, а также с дополнительными переменными, напрямую не контролируемыми фирмами (environmental variables). Применяемый нами бутстрап основан на следующих предположениях: 1) выборка представляет собой независимые одинаково распределенные случайные величины с непрерывной совместной функцией распределения; 2) производственная граница — гладкая; 3) вероятность наблюдения фирм на границе приближается к единице с увеличением размера выборки. Симуляции для многопродуктовой и многофакторной производственный функции показывают почти полное исключение смещения после применения предложенной нами оценки. В завершении статьи методология применяется для оценки эффективности банковского сектора Японии по данным 2013 г. Разработанный на основе этой статьи пакет «rDEA» на языке R доступен в репозитариях GitHub и CRAN.

Ключевые слова: бутстрап, ценовая эффективность, анализ границы, банковское дело.

Статья поступила в редакцию 26.09.18

Статья рекомендована в печать 17.12.18

Контактная информация:

Бесстремянная Галина Евгеньевна — PhD по экономике, канд. экон. наук; gbesstre@nes.ru
Симм Як — PhD in Computer Science; jaak.simm@kuleuven.be

Appendix

Microeconomic framework

The Cobb–Douglas production function for each firm j is taken in the form

$$y_{mj} = A_m \prod_{n=1}^{N} x_{nmj}^{\alpha_{nm}}, \tag{A.1}$$

where x_{nm} is the quantity of the n-th input, used to produce the m-th output, so that $x_n = \sum_{m=1}^{M} x_{nm}$ [Resti, 2000] A_m and α_{nm} are corresponding parameters.

Below we omit index j for simplicity. The derived optimal demand for x_{nm}^{*} becomes a function of outputs and input prices (see [Shephard , 1981; Resti, 2000]):

$$x_{nm}^{*} = \frac{(y_m^{*} / A_m)^{1/\sum_{n=1}^{N}\alpha_{nm}} \alpha_{nm}}{\prod_{n=1}^{N} \alpha_{nm}^{\alpha_{nm}/\sum_{n=1}^{N}\alpha_{nm}}} / \frac{w_n}{\prod_{n=1}^{N} w_n^{\alpha_{nm}/\sum_{n=1}^{N}\alpha_{nm}}} = \frac{(y_m^{*} / A_m)^{1/\rho_m} \alpha_{nm} T_m}{w_n}, \tag{A.2}$$

where $\rho_m = \sum_{n=1}^{N} \alpha_{nm}$ and $T_m = \prod_{n=1}^{N} \left(\frac{w_n}{\alpha_{nm}} \right)^{\alpha_{nm}/\rho_m}$.

Cost inefficiency is added to $(N-1)$ inputs, and then the value of the N-th input is computed, so that the the level of input-oriented efficiency for each firm does not change [Resti, 2000]. Formally,

$$x_{nm} = x_{nm}^{*} \eta_{nm}, n = 1, \ldots, N-1, \eta_{nm} > 0; \tag{A.3}$$

$$x_{Nm} = \left(\frac{y_m^{*}}{A_m \prod_{n=1}^{N-1} x_{nm}^{\alpha_{nm}}} \right)^{1/\alpha_{Nm}}. \tag{A.4}$$

Substituting y_m^{*} in (A.4) by $A_m \prod_{n=1}^{N} x_{nm}^{*}$, and for each $n < N-1$ replacing x_{nm} by $x_{nm}^{*} \eta_{nm}$, we obtain

$$x_{Nm} = \left(\frac{y_m^{*}}{A_m \prod_{n=1}^{N-1} x_{nm}^{\alpha_{nm}}} \right)^{1/\alpha_{Nm}} = x_{Nm}^{*} \prod_{n=1}^{N-1} \eta_{Nm}^{-\alpha_{nm}/\alpha_{Nm}}. \tag{A.5}$$

The cost efficiency is:

$$\gamma = \frac{\boldsymbol{wx^{opt}}}{\boldsymbol{wx}} = \frac{\sum_{n=1}^{N} x_n^{opt} w_n}{\sum_{n=1}^{N} x_n w_n}. \tag{A.6}$$

As regards computing the denominator, the expressions for x_{nm} from (A.3) and (A.5) allow obtaining:

$$x_n w_n = w_n \sum_{m=1}^{M} \overset{*}{x}_{nm} \eta_{nm}, n = 1,..., N-1; \tag{A.7}$$

$$x_N w_N = w_N \sum_{m=1}^{M} \overset{*}{x}_{Nm} \prod_{n=1}^{N-1} \eta_{Nm}^{-\alpha_{nm}/\alpha_{Nm}}. \tag{A.8}$$

Using (A. 2) and (A.4) we express each $\overset{*}{x}_{nm}$ in terms of y_m, so the total cost in a given point (**x, y**,) becomes:

$$\sum_{n=1}^{N} x_n w_n = \sum_{n=1}^{N-1} \sum_{m=1}^{M} (\overset{*}{y}_m / A_m)^{1/\rho_m} \alpha_{nm} T_m \eta_{nm} + \sum_{m=1}^{M} (\overset{*}{y}_m / A_m)^{1/\rho_m} \alpha_{Nm} T_m \prod_{n=1}^{N-1} \eta_{nm}^{-\alpha_{nm}/\alpha_{Nm}}. \tag{A.9}$$

To calculate the nominator of (A.6), we use (A. 2) to express x_n^{opt} in terms of y_m:

$$x_n^{opt} w_n = \sum_{m=1}^{M} (y_m / A_m)^{1/\rho_m} \alpha_{nm} T_m. \tag{A.10}$$

Then,

$$\sum_{n=1}^{N} x_n^{opt} w_n = \sum_{n=1}^{N} \sum_{m=1}^{M} (y_m / A_m)^{1/\rho_m} \alpha_{nm} T_m. \tag{A.11}$$

Finally, since $y_m = \overset{*}{y}_m \theta^{\rho_m}$, we can rewrite

$$\sum_{n=1}^{N} x_n^{opt} w_n = \sum_{n=1}^{N} \sum_{m=1}^{M} (\overset{*}{y}_m / A_m)^{1/\rho_m} \theta \alpha_{nm} T_m. \tag{A.12}$$

Cost efficiency γ is calculated as follows:

$$\gamma = \frac{\boldsymbol{wx^{opt}}}{\boldsymbol{wx}} = \frac{\sum_{n=1}^{N} \sum_{m=1}^{M} (\overset{*}{y}_m / A_m)^{1/\rho_m} \theta \alpha_{nm} T_m}{\sum_{n=1}^{N-1} \sum_{m=1}^{M} (\overset{*}{y}_m / A_m)^{1/\rho_m} \alpha_{nm} T_m \eta_{nm} + \sum_{m=1}^{M} (\overset{*}{y}_m / A_m)^{1/\rho_m} \alpha_{Nm} T_m \prod_{n=1}^{N-1} \eta_{nm}^{-\alpha_{nm}/\alpha_{Nm}}}. \tag{A.13}$$

Our estimations with Japanese data and results from the empirical literature (see [Liu, Ondrich, Ruggiero, 2012; Wang, 2003; Banker, Gadh, Gorr, 1993; Giokas, 1991]) show that input elasticities do not vary appreciably for banking outputs employed in this paper. Therefore, we impose a simplifying assumption $\alpha_{nm} \equiv \alpha_n$, which leads to $T_m \equiv T$ and $\rho_m \equiv \rho$. Accordingly, it becomes reasonable to add inefficiencies to inputs, so that $\eta_{nm} \equiv \eta_n$. The assumptions allow computing cost efficiency γ as follows:

$$\gamma = \frac{T\theta \sum_{n=1}^{N} \alpha_n \sum_{m=1}^{M} \left(\overset{*}{y}_m / A_m\right)^{1/\rho}}{T \sum_{n=1}^{N-1} \alpha_n \eta_n \sum_{m=1}^{M} \left(\overset{*}{y}_m / A_m\right)^{1/\rho} + T\alpha_N \prod_{n=1}^{N-1} \eta_{nm}^{-\alpha_{nm}/\alpha_{Nm}} \sum_{m=1}^{M} \left(\overset{*}{y}_m / A_m\right)^{1/\rho}}. \tag{A.14}$$

Canceling T and $\sum_{m=1}^{M}(y_m^* / A_m)^{1/\rho}$ leads to:

$$\gamma = \frac{\theta \sum_{n=1}^{N} \alpha_n}{\sum_{n=1}^{N-1} \alpha_n \eta_n + \alpha_N \prod_{n=1}^{N-1} \eta_n^{-\alpha_n/\alpha_N}}. \tag{A.15}$$