Centre for
Economic
and Financial
Research
at
New Economic
School

December
2017

Robust estimation of cost efficiency in non-parametric frontier models

Galina Besstremyannaya
Jaak Simm
Sergei Golovan

Working Paper No 244

CEFIR /NES Working Paper series

Robust estimation of cost efficiency in non-parametric frontier models

Galina Besstremyannaya* Jaak Simm** Sergei Golovan***

December 14, 2017

Abstract

The paper proposes a bootstrap methodology for robust estimation of cost efficiency in data envelopment analysis. 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. We consider the cases with absence and presence of environmental variables. Simulation analyses with multi-input multi-output production function demonstrate consistency of the new algorithm in terms of the coverage of the confidence intervals for true cost efficiency.

Finally, we offer real data estimates for Japanese banking industry. Using the nationwide sample of Japanese banks in 2013, we show that the bias of cost efficiency scores may be linked to the bank charter and the presence of the environmental variables in the model.

A package 'rDEA', developed in the R language, is available from the GitHub and CRAN repository.

Keywords: data envelopment analysis, cost efficiency, bias, bootstrap, banking

JEL Classification Codes: C44, C61, G21

*Centre for Economic and Financial Research at New Economic School, gbesstremyannaya@cefir.ru
**University of Leuven
***New Economic School

1 Introduction

Data envelopment analysis (DEA) (Charnes et al. (1978)) is a linear optimization technique, stemming from the seminal work of Farrell (1957), who defined technical and price efficiency of a firm, and proposed a method of constructing a frontier as a linear convex hull surface to envelop observations. The efficiency scores of each firm are estimated according to the distance of the firm from the empirical frontier of efficient firms. However, the empirical frontier may fail to incorporate unobservable but very efficient firms (Simar and Wilson (1998)). So the efficiency scores, which are linked to the empirical frontier, are upward biased. The standard approaches for consistent correction of the bias in 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 semi-parametric bootstrap in presence of additional inputs (so called environmental variables, which are not directly controlled by producers) (Simar and Wilson (2000b); Simar and Wilson (1998); Simar and Wilson (2007)).[1] Concerning cost minimization DEA (Fare et al. (1985)), practitioners suggest a direct modification of the Simar and Wilson (1998) and the Simar and Wilson (2007) bootstrap (de Borger et al. (2008)).

In this paper we show that the direct modification the Simar and Wilson (1998) and the Simar and Wilson (2007) bootstrap is inconsistent and 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 algorithm is applied to bias correction and estimating returns to scale in cost minimization DEA. The results of the simulations for multi-input multi-output Cobb-Douglas production function with correlated outputs, and correlated technical and cost efficiency, show consistency of our proposed algorithm in terms of coverage probability of confidence intervals for true cost efficiency, 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 naive scores, we demonstrate that the direct modification of Simar and Wilson (1998) and Simar and Wilson (2007) bootstrap is consistent.

An application of the algorithm to real data of 106 Japanese banks in the fiscal year 2013 demonstrates re-ranking of banks according to their bias-corrected cost efficiency scores, as well as shows heterogeneity of bias according to bank charter.

The remainder of the paper is structured as follows. Section 2 reviews theoretical framework for bias correction of technical efficiency scores (using an example of input orientation). Section 3 demonstrates inconsistency of a direct application of Simar and Wilson (1998) or Simar and Wilson (2007) bootstrap and offers an alternative bootstrap algorithm for robust estimation of Fare et al. (1985) cost efficiency in absence (presence) of environmental variables. Section 4 conducts simulations for various data generating processes for production frontier and technical and cost inefficiencies. Section 5 provides real data estimates with nationwide sample of Japanese banks. Appendix sets up microeconomic framework for the existence of technical and cost inefficiencies, and gives the theoretical details for the simulations.

Our estimations are conducted with an R package 'rDEA' (Simm and Besstremyannaya (2016)), which is available from GitHub and CRAN repositary.

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

2 Estimates of input-oriented efficiency

2.1 Naive score

Denote the existing technology, which produces outputs $\mathbf{y}_m$ $(m = 1, ..., M)$ using inputs $\mathbf{x}_n$ $(n = 1, ..., N)$ as $T = \{(\mathbf{x}, \mathbf{y}) : \mathbf{x} \geq \mathbf{X}\boldsymbol{\lambda}, \mathbf{y} \leq \mathbf{Y}\boldsymbol{\lambda}, \boldsymbol{\lambda} \geq \mathbf{0}\}$. Input set $L(\mathbf{y})$ (Coelli et al. (1994); Shephard (1981)) contains inputs, that can produce a given amount of output under T, so that $L(\mathbf{y}) = \{(\mathbf{x}) : (\mathbf{x}, \mathbf{y}) \in T\}$. The important assumptions 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 $(j = 1, ..., J)$ is defined as a solution to the below optimization problem (for constant returns to scale, CRS, Charnes et al. (1978)):

$$
\begin{aligned}
&\min_{\theta_j, \boldsymbol{\lambda}} \theta_j \\
\text{s.t.} \quad & -y_{mj} + \sum_{i=1}^{J} \lambda_i y_{mi} \geq 0, \quad m = 1, ..., M, \\
& \theta_j x_{nj} - \sum_{i=1}^{J} \lambda_i x_{ni} \geq 0, \quad n = 1, ..., N, \\
& \lambda_i \geq 0, \quad i = 1, ..., J.
\end{aligned}
\tag{1}
$$

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

It should be noted that the 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} (m = 1, . . . , M). Similarly, there are N constraints on each input $x_n j$.

2.2 Bias correction

The estimates of input-oriented efficiency are upwards biased, since the estimated boundary $\widehat{L}^{\partial}(\mathbf{y})$ of the input set is based on the sample of the observed $DMUs$, which may fail to incorporate the most efficient $DMUs$ in the true $L(\mathbf{y})$ (Simar and Wilson (1998); Simar and Wilson (2000a)). Therefore, the bootstrap methods correct for the bias, constructing pseudo-samples which would belong to $\widehat{L}(\mathbf{y})$. Then, according to the re-centering idea of bootstrap, for each DMU i $\text{bias}\, \theta_i = E(\hat{\theta}_i) - \theta_i = \widehat{\text{bias}\, \hat{\theta}_i} = \text{bias}\, \hat{\theta}_i^* = E(\hat{\theta}_i^*) - \hat{\theta}_i$. In particular, the homogeneous smoothed bootstrap projects each observation on the frontier and then "pushes" it inside the $\widehat{L}(\mathbf{y})$ (Simar and Wilson (2008); Simar and Wilson (1998)):

1. Estimate naive scores $\hat{\theta}_1, ..., \hat{\theta}_J$, for each $i = 1, ..., J$ according to system (1). Assume $(\theta_1, ..., \theta_J)$ are i.i.d. with pdf $f(\cdot)$.

2. Loop B times to obtain J sets of bootstrap estimates $\{\hat{\theta}_{ib}^*\}_{b=1}^{B}$.

   2.1 Obtain a smooth estimate $\hat{f}(\theta)$ and for each $i = 1, ..., J$ draw θ_{ib}^* from this estimate.[2]

   2.2 Assume homogeneous distribution of joint density of θ in input-output space,

   i.e. $\hat{f}(\theta_i | (\mathbf{x}_i, \mathbf{y}_i)) = \hat{f}(\theta_i)$ and assign $\mathbf{x}_{ib}^* = \dfrac{\hat{\theta}_i}{\theta_{ib}^*} \mathbf{x}_i$.

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

2.3 Calculate $\hat{\theta}^*_{ib}$ for $(\mathbf{x}^*_{ib}, \mathbf{y}_i)$.

3. $\widehat{\text{bias}\,\theta_i} = \dfrac{1}{B}\displaystyle\sum_{b=1}^{B} \hat{\theta}^*_{ib} - \hat{\theta}_i$ and bias-corrected score $\hat{\hat{\theta}}_i = \hat{\theta}_i - \widehat{\text{bias}\,\hat{\theta}_i}$.

Rescaling at step (2.2) guarantees that pseudo-samples $\{(\mathbf{x}^*_{ib}, \mathbf{y}_i)\}_{b=1}^{B} \in \widehat{L}(\mathbf{y})$. Indeed, input-oriented efficiency evaluates the potential of $DMU\ i$ for maximal reduction of inputs, holding the amount of outputs constant. The constraints $\mathbf{x}_i \geq \mathbf{X}\boldsymbol{\lambda}$ imply inputs are larger than possible. Therefore, multiplications of each input by $\hat{\theta}_i$, $0 \leq \hat{\theta}_i \leq 1$, projects it to $\widehat{L}^{\partial}(\mathbf{y})$, so that the projected observation become an estimate of an efficient input level with coordinates $(\hat{\theta}\mathbf{x}_i, \mathbf{y}_i)$. 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 $\widehat{L}(\mathbf{y})$.

In presence of an r-dimensional vector of *environmental* variables $\mathbf{z}$ (i.e. a special type of inputs that are not directly controlled by producers) Simar and Wilson (2007) propose semi-parametric bootstrap for correcting the bias of distance function score δ, the reciprocal of θ.[3] The algorithm, in 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 $\mathbf{x}$ does not depend on $\mathbf{z}$ (Simar and Wilson (2011b)).

1. Estimate naive 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 $\delta_i = \mathbf{z}_i\boldsymbol{\beta} + \varepsilon_i \geq 1$, where ε_i are i.i.d. and independent from $\mathbf{z}_i$, $\varepsilon_i \sim N(0, \sigma^2_{\varepsilon})$ with left truncation at $(1 - \mathbf{z}_i\boldsymbol{\beta})$.

2. Use observations for which $\hat{\delta} > 1$ to obtain $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}_{\varepsilon}$ in the truncated regression $\hat{\delta}_i = \mathbf{z}_i\boldsymbol{\beta} + \varepsilon_i \geq 1$.

3. Loop B times to obtain J sets of bootstrap estimates $\{\hat{\delta}^*_{ib}\}_{b=1}^{B}$.

   3.1 For each $i = 1, ..., J$ draw ε_i from $N(0, \hat{\sigma}^2_{\varepsilon})$ with left truncation at $(1 - \mathbf{z}_i\hat{\boldsymbol{\beta}})$.

   3.2 For each $i = 1, ..., J$ compute $\delta^*_i = \mathbf{z}_i\hat{\boldsymbol{\beta}} + \varepsilon_i$.

   3.3 Assign $\mathbf{x}^*_{ib} = \dfrac{\delta^*_{ib}}{\hat{\delta}_i}\mathbf{x}_i$.

   3.4 Calculate $\hat{\delta}^*_{ib}$ for $(\mathbf{x}^*_{ib}, \mathbf{y}_i)$.

4. $\widehat{\text{bias}\,\delta_i} = \dfrac{1}{B}\displaystyle\sum_{b=1}^{B} \hat{\delta}^*_{ib} - \hat{\delta}_i$ and bias-corrected score $\hat{\hat{\delta}} = \hat{\delta} - \widehat{\text{bias}\,\hat{\delta}}$.

2.3 Returns to scale

The above algorithm consistently estimates the sampling distribution of the original efficiency scores and therefore, is applicable for testing returns to scale (Simar and Wilson (2002)). For instance, the null hypothesis of constant returns to scare verses an alternative hypothesis of variable returns to scale may be tested through bootstrapping an appropriate test statistics under the null hypothesis (Simar and Wilson (2011b), Simar and Wilson (2008), Simar and Wilson (2002)). The simulation analyses show that statistics equal to the ratio of mean scores $\sum_{j=1}^{J} \theta^{CRS}(\mathbf{x}_i, \mathbf{y}_i) \Big/ \sum_{j=1}^{J} \theta^{VRS}(\mathbf{x}_i, \mathbf{y}_i)$ provides for tests of good power (Simar

[3] θ, which is bounded between 0 and 1, could not be used for computational reasons in estimating truncated regression (Simar and Wilson (2008)).

and Wilson (2002)). Yet, the most appropriate test statistics, stemming from the theoretical result in Kneip et al. (2008) is the mean of ratios $\frac{1}{J}\sum_{j=1}^{J}\theta^{CRS}(\mathbf{x}_i,\mathbf{y}_i)/\theta^{VRS}(\mathbf{x}_i,\mathbf{y}_i)$ (Simar and Wilson (2011b)).

3 Estimates of cost efficiency

3.1 Naive score with given input prices

Denote $\mathbf{w}_j$ the vector of input prices. Fare et al. (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):

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

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

3.2 Proposed bootstrap algorithm for the Fare et al. (1985) cost efficiency

Similarly to input-oriented efficiency scores, Fare et al. (1985) cost efficiency scores are linked to $\widehat{L}^{\partial}(\mathbf{y})$ and therefore, are upwards-biased. Yet, a direct modification of the Simar and Wilson (1998) or the Simar and Wilson (2007) algorithm to bias correction of cost efficiency score γ, which simply replaces θ by γ at steps 2.2 (step 3.3) (de Borger et al. (2008)), is inconsistent. Indeed, let's look at a given observation i with coordinates $\mathbf{x}_i$ (point P at Figure (2)). By definition of input-oriented efficiency, point P''', which is an intersection of the ray from the origin to P and $\widehat{L}^{\partial}(\mathbf{y})$, has coordinates $\hat{\theta}\mathbf{x}_i$. The hyperplane, set by the cost function $\mathbf{w}_i\mathbf{x}_i$ and tangent to $\widehat{L}^{\partial}(\mathbf{y})$, 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 $P'' \notin [P''',P]$, i.e. $P'' \in [P',P'']$. So the vector of bootstrapped inputs, obtained at step 2.2 of a direct modification of Simar and Wilson (1998) algorithm, may be outside the $\widehat{L}(\mathbf{y})$. (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 $L(\mathbf{y})$ and free disposability of inputs are importantly exploited in our argument.

To correct for the bias of the Fare et al. (1985) cost efficiency we propose the following bootstrap, which 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

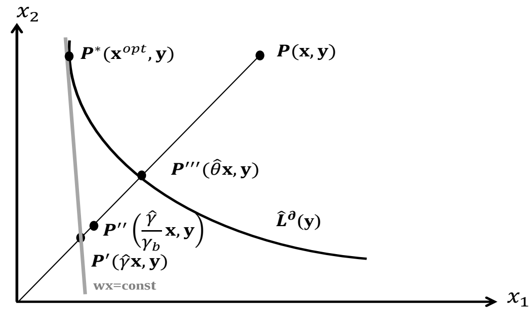

Figure 1: **Bias correction of the Fare et al. (1985) cost efficiency, isoquant in the two-input space**

bootstrapped inputs are "pushed" inside the $\widehat{L}(\mathbf{y})$. 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 $(\gamma_1, ..., \gamma_J)$ 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 $x^*_{ib} = \frac{\hat{\theta}_i}{\theta^*_{ib}} x_i (b = 1, ..., B)$.

4. Calculate $\hat{\gamma}^*_{ib}$ for (x^*_{ib}, y_i).

5. For each i, bias $\hat{\gamma}_i = \frac{1}{B} \sum_{b=1}^{B} \hat{\gamma}^*_{ib} - \hat{\gamma}_i$.

In presence of the environmental variables, given Simar and Wilson (2007) assumption about separability of $\mathbf{x}$ and $\mathbf{z}$ (i.e. the fact that $L^{\partial}(\mathbf{y})$ does not depend on $\mathbf{z}$), we propose the following algorithm for the reciprocal of Fare et al. (1985) cost efficiency score, denoted δ^γ_i:

1. Estimate reciprocals of naive cost efficiency scores $\hat{\delta}^\gamma_1, ..., \hat{\delta}^\gamma_J$, for each $i = 1, ..., J$ using system (3). Assume $\delta^\gamma_i = \mathbf{z}^\gamma_i \boldsymbol{\beta}^\gamma + \psi_i \geq 1$, where ψ_i are i.i.d. and independent from $\mathbf{z}^\gamma_i$, $\psi_i \sim N(0, \sigma^2_\psi)$ with left truncation at $(1 - \mathbf{z}^\gamma_i \boldsymbol{\beta}^\gamma)$.

2. Estimate naive input-oriented 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 $\delta_i = \mathbf{z}_i \boldsymbol{\beta} + \varepsilon_i \geq 1$, where ε_i are i.i.d. and independent from $\mathbf{z}_i$, $\varepsilon_i \sim N(0, \sigma^2_\varepsilon)$ with left truncation at $(1 - \mathbf{z}_i \boldsymbol{\beta})$.

3. Use observations for which $\hat{\delta} > 1$ to obtain $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}_\varepsilon$ in the truncated regression $\hat{\delta}_i = \mathbf{z}_i \boldsymbol{\beta} + \varepsilon_i \geq 1$

4. Loop B times to obtain J sets of bootstrap estimates $\{\hat{\delta}^*_{ib}\}$, $b = 1, ..., B$.

   4.1 For each $i = 1, ..., J$ draw ε_i from $N(0, \hat{\sigma}^2_\varepsilon)$ with left truncation at $(1 - \mathbf{z}_i \hat{\boldsymbol{\beta}})$.

   4.2 For each $i = 1, ..., J$ compute $\delta^*_i = \mathbf{z}_i \hat{\boldsymbol{\beta}} + \varepsilon_i$.

   4.3 Given the semi-parametric dependence of δ on $\mathbf{z}$, assign $\mathbf{x}^*_{ib} = \dfrac{\delta^*_{ib}}{\hat{\delta}_i} \mathbf{x}_i$.

   4.4 Calculate $\hat{\delta}^{\gamma *}_{ib}$ for $(\mathbf{x}^*_{ib}, \mathbf{y}_i)$.

5. Owing to semi-parametric dependence of δ^γ on $\mathbf{z}^\gamma$, we can compute $\widehat{\text{bias}\,\delta_i^\gamma} = \dfrac{1}{B}\sum_{b=1}^{B} \hat{\delta}_{ib}^{\gamma*} - \hat{\delta}_i^\gamma$ and $\hat{\hat{\delta}}^\gamma = \hat{\delta}^\gamma - \widehat{\text{bias}\,\hat{\delta}^\gamma}$

Note that $\{\mathbf{z}_i\} \subset \{\mathbf{z}_i^\gamma\}$. 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.

3.3 Proposed returns to scale test in Fare et al.'s (1985) cost minimization DEA

Since our proposed bootstrap algorithm consistently estimates the sampling distribution of the original cost efficiency scores under correctly specified returns to scale, it may be applicable for testing returns to scale for the production possibility frontier in cost minimization DEA.

Namely, in each bootstrap loop we first, conduct estimates with input-oriented efficiency under the null hypothesis and rescale inputs. Second, we compute cost efficiency scores δ^γ for rescaled inputs under the null and alternative hypotheses and get the values of the test statistics $\dfrac{1}{J}\sum_{j=1}^{J} \delta^{\gamma,VRS}(\mathbf{x}_i, \mathbf{y}_i)/\delta^{\gamma,CRS}(\mathbf{x}_i, \mathbf{y}_i)$ (Simar and Wilson (2011b)).

Note that our cost-minimization procedure relies on an input-oriented model. In other words, the necessary condition for the presence of constant returns to scale in the cost-minimization model is the non-rejection of the CRS hypothesis both in the RTS test for an input-oriented model and for cost-minimization model.

3.4 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 $\bar{\mathbf{x}}_j = (w_{1j}x_{1j}, ..., w_{Nj}x_{Nj})^T$, $\bar{\mathbf{X}} = (\bar{\mathbf{x}}_1, ..., \bar{\mathbf{x}}_J)^T$, where $\mathbf{w}_j$ is a vector of prices for each input $\mathbf{x}_j$. "New" cost efficiency for DMU j is defined as

$$\bar{\gamma}_j = \mathbf{e}\bar{\mathbf{x}}_j^{opt}/\mathbf{e}\bar{\mathbf{x}}_j \tag{4}$$

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

$$\begin{aligned} &\min_{\bar{\mathbf{x}}_j, \boldsymbol{\lambda}} \mathbf{e}\bar{\mathbf{x}}_j \\ \text{s.t.} \quad & -y_{mj} + \sum_{i=1}^{J} \lambda_i y_{mi} \geq 0, \quad m = 1, ..., M, \\ & \bar{x}_{nj} - \sum_{i=1}^{J} \lambda_i \bar{x}_{ni} \geq 0, \quad n = 1, ..., N, \\ & \lambda_i \geq 0, \quad i = 1, ..., J. \end{aligned} \tag{5}$$

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

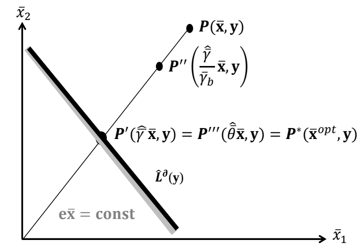

Figure 2: **Bias correction of Tone (2002) cost efficiency, isoquant in two-input space**

3.5 Proposed bootstrap algorithm for the Tone (2002) new cost efficiency

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

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

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 the below aggregate constraint:

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

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 $P' = P'''$. In other words, as is noted in Tone (2002) (theorem 6), the "new" cost efficiency point is also "new" technically efficient.[4] So a consistent bias correction of Tone (2002) "new" cost efficiency score may be conducted through a direct application of Simar and Wilson (1998) (Simar and Wilson (2007)) algorithm, so that the following rescaling is implemented at step (3) (step (3.3)): $\bar{x}^*_{i,b} = \dfrac{\hat{\bar{\gamma}}}{\bar{\gamma}^*_{i,b}}\bar{x}_i$. Indeed, as $L_n^{\partial}(\mathbf{y})$ is set by the aggregate constraint (7), $P'' \in [P', P]$ is equivalent to $P'' \in [P''', P]$. Therefore, rescaling guarantees that each component of $\bar{\mathbf{x}}_b$ is larger than the corresponding component of the original vector $\bar{\mathbf{x}}$, and vector $\bar{\mathbf{x}}_b$ lies in the necessary subspace relative to $L_n^{\partial}(\mathbf{y})$ (Besstremyannaya (2013)).

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

4 Simulations

4.1 Microeconomic framework

The Cobb-Douglas production function, commonly used in the non-parametric efficiency analysis in the banking industry (Kneip et al. (2011); Fethi and Pasiouras (2010); Thanassoulis et al. (2008); Kneip et al. (2008); Badin and Simar (2003); Simar and Wilson (2002); Simar and Wilson (2000b); Kittelsen (1999); Banker et al. (1993)) is taken in the form (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 distributions, where vectors of means and variance-covariance matrices 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 as a multi-output industry, therefore, we exploit a two-output and three-input models.

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

In absence of environmental variables, inefficiencies are added so that $\mathbf{y} = \mathbf{y}^*\theta^\rho$, $0 < \theta^\rho \leq 1$ (Kneip et al. (2011); Badin and Simar (2003); Simar and Wilson (2002); Simar and Wilson (2000b); Resti (2000); Kittelsen (1999)). Then, owing to homothetic property of cost function, the input-oriented efficiency is θ.

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 $n = 1, ..., N-1$; $\eta_{nm} > 0$. Then, cost efficiency γ is calculated as follows (Appendix, eq.A.13):

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

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 the results in the empirical literature (Liu et al. (2012); Wang (2003); Banker et al. (1993); Giokas (1991)) show that input elasticities do not vary appreciably for banking outputs, employed in this paper. So under $\alpha_{nm} \equiv \alpha_n$ and $\eta_{nm} \equiv \eta_n$ we obtain (eq.A.15):

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

We use constant returns to scale with $(\alpha_1, \alpha_2, \alpha_3) = (0.05, 0.05, 0.9)$. Cost inefficiencies are added to x_2 and x_3 and analytically computed for x_1. The dimensions of the problem are crucial analyzing the coverage

[5]As robustness check we conducted a second set of simulations with the data coming from the intermediation approach and found similar results in terms of coverage probabilities.

probabilities of confidence intervals to $(1-\alpha)$ in bias-correction without environmental variables (e.g. our results may be contrasted to similar findings for technical efficiency scores, reported in Kneip et al. (2008), Table 2)).

Input-oriented efficiency $\theta = 1/(1+\zeta)$, 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 unity.[6]

In presence of environmental variables, we introduce inefficiencies as $\mathbf{y} = \mathbf{y}^*\delta^{-\rho}$, $0 < \delta^{-\rho} \leq 1$, where δ can be expressed as $\mathbf{z}\boldsymbol{\beta}+\varepsilon$. 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 $(1-\mathbf{z}\boldsymbol{\beta})$, $\delta = \mathbf{z}\boldsymbol{\beta}+\varepsilon$. Then eq.(10) modifies to

$$\delta^\gamma(\mathbf{z}) = \frac{\delta}{\rho}\bigg(\sum_{n=1}^{N-1}\alpha_n\eta_n + \alpha_N\prod_{n=1}^{N-1}\eta_n^{-(\alpha_n/\alpha_N)}\bigg) \tag{11}$$

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 and Wilson (2011b) we use 1000 trials with B=2000 iterations on each trial and samples $J = \{50, 100, 200, 300, 400, 600, 800, 1000\}$. For each $\alpha \in \{0.01, 0.05, 0.1\}$ we estimate probabilities of symmetric $(1-\alpha)$ confidence intervals to cover true values of cost efficiency γ $(1/\delta^\gamma)$ in absence (presence) of environmental variables.

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 $E\theta$. So the coordinates of a point on the frontier are $\Big(\mathbf{x}^*\big([E\theta]^\rho\boldsymbol{\mu}_\mathbf{y}, \boldsymbol{\mu}_\mathbf{w}\big), [E\theta]^\rho\boldsymbol{\mu}_\mathbf{y}\Big)$, 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).

4.2 Results

Owing to potential problems of ignoring zero bound in implementing the Silverman (1986) reflection method with the input-oriented efficiency scores θ (Simar and 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 and Wilson (2008)). Since the choice of bandwidth may influence coverage probabilities for small samples (Simar and Wilson (2000b); Kneip et al. (2008)), the simulations in this paper exploit two types of bandwidths: 1) Silverman's (1986) bandwidth for standard normal density function; 2) bandwidth, estimated with least-squares cross-validation and adjusted for sample size (Simar and Wilson (2008)).

The rule of thumb bandwidth, proportional to $J^{-1/(3(M+N+1))}$ in case of bootstrapping θ (Kneip et al. (2008)), is not exploited in our estimations for a few reasons. Firstly, it requires a choice of a factor of proportionality, which may be an additional research task in the analysis with the reciprocal of θ. Secondly, it

[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.

gives comparable results with cross-validation bandwidth for consistent bias correction of technical efficiency scores (Simar and Wilson (1998); Badin and Simar (2003); Simar and Wilson (2000b); Kneip et al. (2008)), and our simulations within cost-minimization framework in terms of θ show similar results on the coverage probabilities for both bandwidths.

In absence of environmental variables (Figure 3) we discover that for a given type of bandwidth and given values α and sample size J, coverage probability of confidence intervals is higher for smaller cost inefficiency (in terms of σ_ν). Cross-validation bandwidth gives coverage probabilities that do not depend on sample size and are in the range of (0.65, 0.91) for $\alpha < 0.1$ and $dim(x) + dim(y) = 5$. Silverman's (1986) bandwidth provides for the worst results, proving inapplicability of normal reference rule. As for the simulation in presence of enviromental variables, where estimation does not involve the use of bandwidths, coverage probability of confidence interval are higher and close to $1-\alpha$ with $J > 600$. 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 $J > 600$.

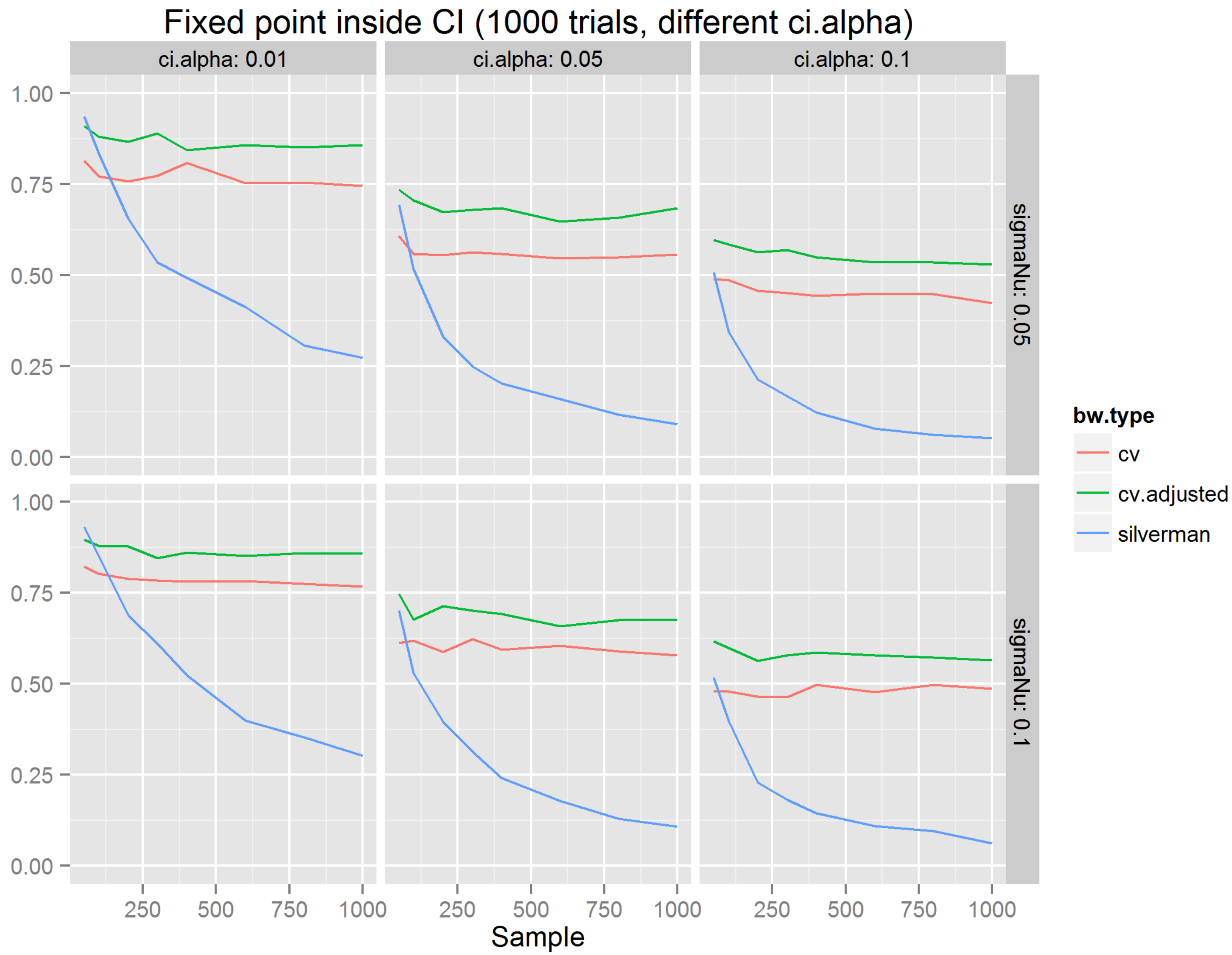

Figure 3: **Coverage probability of confidence intervals for the fixed point,** $dim(x) = 3, dim(y) = 2$

Table 1: **Coverage probability of confidence intervals for homogeneous smooth bootstrap in absence of environmental variables, with sample adjusted cross-validation bandwidth**

		$dim(x) = 3, dim(y) = 2$		
J	σ_ν	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
50	0.050	0.910	0.735	0.597
100	0.050	0.880	0.706	0.584
200	0.050	0.867	0.674	0.563
300	0.050	0.889	0.679	0.569
400	0.050	0.844	0.684	0.550
600	0.050	0.857	0.648	0.536
800	0.050	0.851	0.658	0.536
1,000	0.050	0.858	0.684	0.529
50	0.100	0.896	0.748	0.617
100	0.100	0.879	0.677	0.598
200	0.100	0.878	0.714	0.563
300	0.100	0.845	0.701	0.578
400	0.100	0.861	0.692	0.586
600	0.100	0.852	0.659	0.579
800	0.100	0.860	0.676	0.572
1,000	0.100	0.858	0.675	0.565

Table 2: **Absolute difference between the true and estimated cost efficiency for homogeneous smooth bootstrap in absence of environmental variables, with sample adjusted cross-validation bandwidth**

		$dim(x) = 3, dim(y) = 2$					
J	σ_ν	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.10$	
50	0.050	0.036	[0.012]	0.038	[0.012]	0.037	[0.012]
100	0.050	0.026	[0.008]	0.026	[0.008]	0.026	[0.008]
200	0.050	0.018	[0.005]	0.018	[0.005]	0.018	[0.005]
300	0.050	0.014	[0.004]	0.014	[0.004]	0.014	[0.004]
400	0.050	0.012	[0.003]	0.012	[0.003]	0.012	[0.003]
600	0.050	0.010	[0.002]	0.010	[0.002]	0.010	[0.002]
800	0.050	0.008	[0.002]	0.009	[0.002]	0.008	[0.002]
1,000	0.050	0.007	[0.002]	0.007	[0.002]	0.008	[0.002]
50	0.100	0.037	[0.012]	0.037	[0.012]	0.037	[0.012]
100	0.100	0.025	[0.008]	0.026	[0.008]	0.025	[0.008]
200	0.100	0.018	[0.005]	0.018	[0.005]	0.018	[0.005]
300	0.100	0.014	[0.004]	0.014	[0.004]	0.014	[0.004]
400	0.100	0.012	[0.003]	0.012	[0.003]	0.012	[0.003]
600	0.100	0.010	[0.002]	0.010	[0.002]	0.010	[0.002]
800	0.100	0.008	[0.002]	0.008	[0.002]	0.008	[0.002]
1,000	0.100	0.007	[0.002]	0.008	[0.002]	0.007	[0.002]

Note: Standard deviation in brackets.

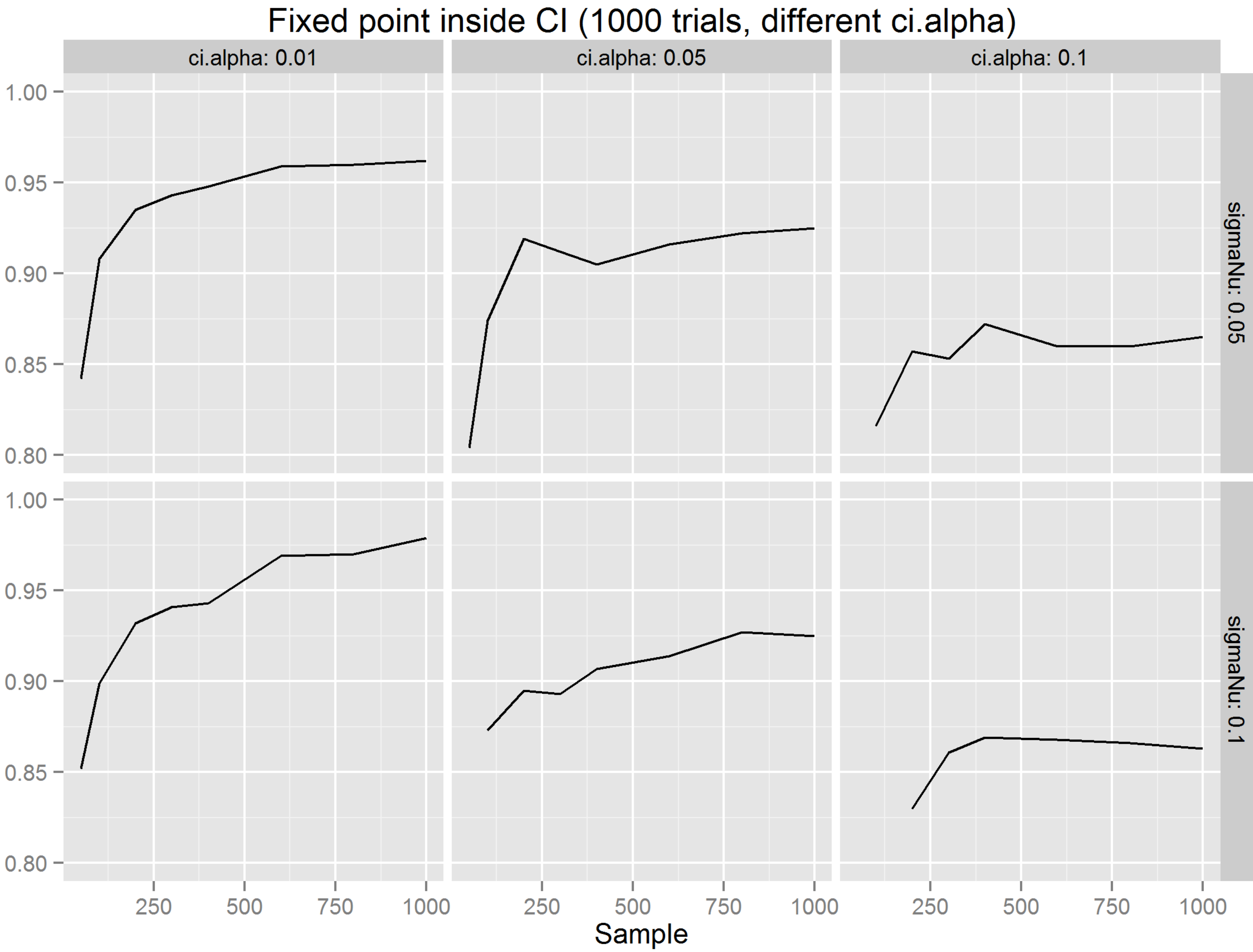

Figure 4: **Coverage probability of confidence intervals for the fixed point in presence of environmental variables**, $dim(x) = 3, dim(y) = 2$, *Case 1*

Table 3: **Coverage probability of confidence intervals for semi-parametric bootstrap in presence of environmental variables,** $dim(x) = 3, dim(y) = 2$

J	σ_ν	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
50	0.050	0.842	0.804	0.794
100	0.050	0.908	0.874	0.816
200	0.050	0.935	0.919	0.857
300	0.050	0.943	0.912	0.853
400	0.050	0.948	0.905	0.872
600	0.050	0.959	0.916	0.860
800	0.050	0.960	0.922	0.860
1,000	0.050	0.962	0.925	0.865
50	0.100	0.852	0.794	0.780
100	0.100	0.899	0.873	0.799
200	0.100	0.932	0.895	0.830
300	0.100	0.941	0.893	0.861
400	0.100	0.943	0.907	0.869
600	0.100	0.969	0.914	0.868
800	0.100	0.970	0.927	0.866
1,000	0.100	0.979	0.925	0.863

Table 4: **Absolute difference between the true and estimated cost efficiency for semi-parametric bootstrap in presence of environmental variables,** $dim(x) = 3, dim(y) = 2$

J	σ_ν	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.10$	
50	0.050	0.043	[0.017]	0.043	[0.017]	0.042	[0.016]
100	0.050	0.029	[0.010]	0.030	[0.010]	0.029	[0.010]
200	0.050	0.020	[0.006]	0.020	[0.005]	0.020	[0.005]
300	0.050	0.016	[0.004]	0.016	[0.004]	0.016	[0.004]
400	0.050	0.014	[0.003]	0.014	[0.003]	0.014	[0.003]
600	0.050	0.011	[0.002]	0.011	[0.002]	0.011	[0.002]
800	0.050	0.010	[0.002]	0.010	[0.002]	0.010	[0.002]
1,000	0.050	0.009	[0.002]	0.009	[0.002]	0.009	[0.002]
50	0.100	0.043	[0.018]	0.042	[0.017]	0.042	[0.016]
100	0.100	0.029	[0.010]	0.030	[0.010]	0.030	[0.010]
200	0.100	0.021	[0.006]	0.021	[0.006]	0.021	[0.006]
300	0.100	0.017	[0.004]	0.017	[0.004]	0.017	[0.004]
400	0.100	0.014	[0.004]	0.014	[0.003]	0.014	[0.003]
600	0.100	0.012	[0.003]	0.012	[0.003]	0.012	[0.003]
800	0.100	0.010	[0.002]	0.010	[0.002]	0.010	[0.002]
1,000	0.100	0.009	[0.002]	0.009	[0.002]	0.009	[0.002]

Note: Standard deviation in brackets.

5 Efficiency estimates for Japanese banks

5.1 Data

We use the 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 charter 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 and Yoshida (1996); Fukuyama and Weber (2002); Barros et al. (2012)) or revenue from loans and revenue from other business activities (intermediation approach, e.g. Kasuya (1986); Fukuyama (1993); Fukuyama (1995); Takahashi (2000); Fukuyama and Weber (2010)) (Thanassoulis et al. (2008); Tortosa-Austina (2002)). In each model the inputs are labor (total employees), capital (premises, real estate and intangibles) and funds from customers (Kasuya (1986); Kasuya (1989); Fukuyama (1993); Fukuyama (1995); Hori and Yoshida (1996); McKillop et al. (1996); Glass et al. (1998); Fukuyama and Weber (2002); Miyakoshi and Tsukuda (2004); Fukuyama and Weber (2008); Barros et al. (2012)). The proxies for input prices are, respectively, personnel expenditure/total employees, capital expenditure/capital and fund-raising expenditure/funds from customers (Kasuya (1986); Kasuya (1989); McKillop et al. (1996); Fukuyama and 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 (Aly et al. (1990); Simar and Wilson (2007)), ratio of loan loss provisions to total loans (Altunbas et al. (2000), Drake and Hall (2003), Drake et al. (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 and 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 5).

The most recent data for gross regional product is available for the fiscal year which runs from Apr 2013 to Mar 2014, so we take the sample of Japanese banks in this year. Our sample represents the whole banking industry in Japan, yet, its size is only 106. However, the results of our simulations demonstrate high coverage probabilities in case of cross-validation bandwidth even for such small samples.

[7] Note that intermediation approach prevails in international literature (Fethi and Pasiouras (2010)), yet, asset approach is more spread in the analyses on 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.

Table 5: **Descriptive statistics in the fiscal year 2013**

Variable	Definition	Obs	Mean	Std.Dev.	Min.	Max.
Inputs						
x_1	labor = total employees (including board)	106	2714	4481	312	31461
x_2	capital = premises and real estate + intangibles	106	26	123	.044	1125
x_3	funds from customers = total deposits + negotiable certificates of deposits +call money + bills sold + borrowed money + foreign exchange deposits + other deposits	106	7366	21589	219	157288
Outputs						
Asset approach						
y_1	performing loans = total loans − nonperforming loans	106	4614	12566	158	89543
y_2	securities and other interest bearing assets	106	2575	8434	1.4	66543
Intermediation approach						
y_3	revenue from loans = interest on loans and discounts + interest on bills bought	106	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	106	63	214	.706	1428
Input prices						
w_1	labor price = (general and administrative expenses-depreciation)/total employees	106	.017	.007	.009	.065
w_2	capital price = (expenditure on premises and fixed assets)/x_2	106	1.126	1.399	.136	10.442
w_3	price of funds = fund raising expenditure/x_3	106	.001	.001	0	.005
Bank variables						
z_1	= ln(branches)	106	4.56	.62	3.05	6.72
z_2	Herfindahl index of product diversity	106	.72	.23	.26	1.44
z_3	nonperforming loan ratio = nonperforming loans /total loans	106	.03	.01	.004	.07
z_4	= 1 if city bank	106	0.05	0.21	0	1
z_5	= 1 if regional bank	106	0.56	0.50	0	1
z_6	= 1 if regional tier 2 (former Sogo) bank	106	0.35	0.48	0	1
z_7	= 1 if trust bank	106	0.03	0.17	0	1
z_8	= 1 if longterm credit bank	106	0.02	0.14	0	1
Prefectural variables						
z_9	rate of growth of gross regional product (in 2010 real terms)	106	1.00	.05	.93	1.18
z_{10}	share of monetary aggregate (M2+ negotiable certificates of deposit) in regional product	106	.70	.32	.40	1.52
z_{11}	share of loans in gross regional product	106	.75	.46	.36	1.95
z_{12}	rate of growth of price of commercial land (in 2010 real terms)	106	.58	.14	.41	.92

Note: Financial variables are in billion yen.

5.2 Results

Estimations are conducted under variable returns to scale with B=2000. As rule of thumb bandwidth may be unstable with moderate samples in estimations without environmental variables, we exploit least squares cross-validation in the choice of bandwidth. We 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 6 shows the estimates of "naive" score $\hat{\gamma}$ ($1/\hat{\delta}^{\gamma}$) and bias-corrected score $\hat{\hat{\gamma}}$ ($1/\hat{\hat{\delta}}^{\gamma}$) for the models, corresponding to asset approach and intermediation approach. In each model mean bias-corrected score is lower than mean "naive" score, while standard deviation of "naive" and bias-corrected scores are close. Bias-corrected score is "to the left" (if compared to the range of "naive" 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 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.

Table 6: **Cost efficiency scores**

Score		**Asset approach**	**Intermediation approach**
$\hat{\gamma}$	mean	0.7189	0.6679
	st.dev.	0.1406	0.1660
	range	[0.4496, 1]	[0.3781, 1]
$\hat{\hat{\gamma}}$	mean	0.6622	0.5906
	st.dev.	0.1300	0.1370
	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	0.6150
	st.dev.	0.1187	0.1205
	range	[0.4832, 0.9732]	[0.3739, 0.8928]

Quantile-quantile plots for $\hat{\hat{\gamma}}$ and $\hat{\gamma}$ ($1/\hat{\delta}^{\gamma}$ and $1/\hat{\hat{\delta}}^{\gamma}$) allow visualizing the bias and its heterogeneity over observations. As may be inferred from Figures $5-6$ the upward bias of $\hat{\gamma}$ ($1/\hat{\delta}^{\gamma}$) does not vary appreciably with bank charter for cost efficiency score under asset approach. However, the heterogeneity depends on bank charter in the model with intermediation approach: the distance from the 45 degree line is the largest for national banks and longterm credit/trust banks. The bias and heterogeneity is larger in presence of environmental variables.

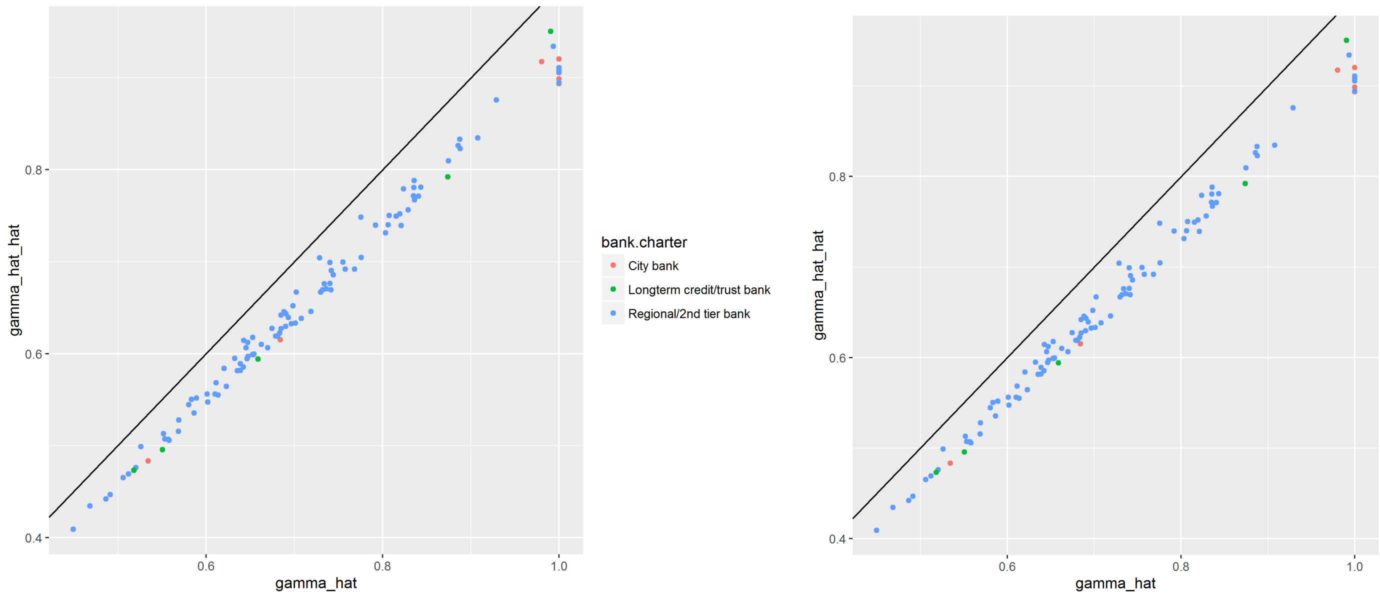

Figure 5: **Quantile-quantile plots for models with asset approach (left) and intermediation approach (right) in absence of environmental variables**

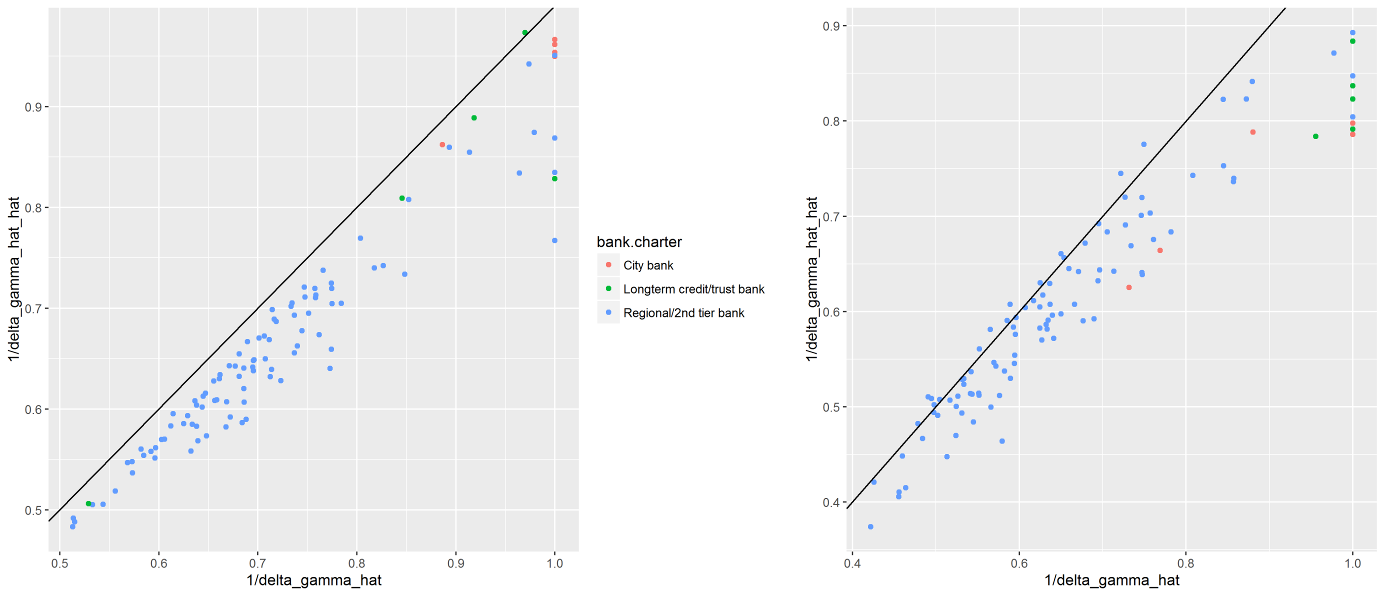

Figure 6: **Quantile-quantile plots for models with asset approach (left) and intermediation approach (right) in presence of environmental variables**

Figure 7 demonstrates re-ranking of banks according to their bias-corrected cost efficiency scores. Indeed, ordered according to monotonically increasing "naive" cost efficiency scores $\hat{\gamma}$ (*green line*), banks have non-monotonic bias-corrected cost efficiency scores $\hat{\hat{\gamma}}$ (*blue line*).

Similarly, Figure 8 indicates re-ranking of banks according to their bias-corrected distance function scores $\hat{\delta}^{\gamma}$.

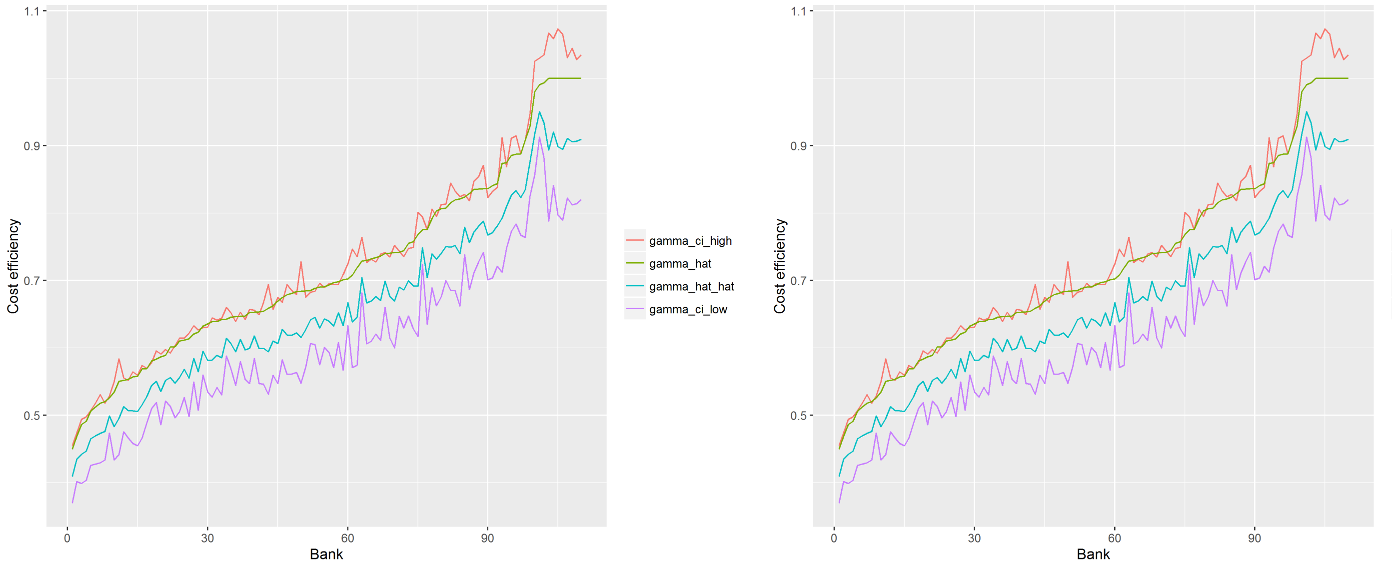

Figure 7: **"Naive" and bias-corrected cost efficiency for models with asset approach (left) and intermediation approach (right) in absence of environmental variables**

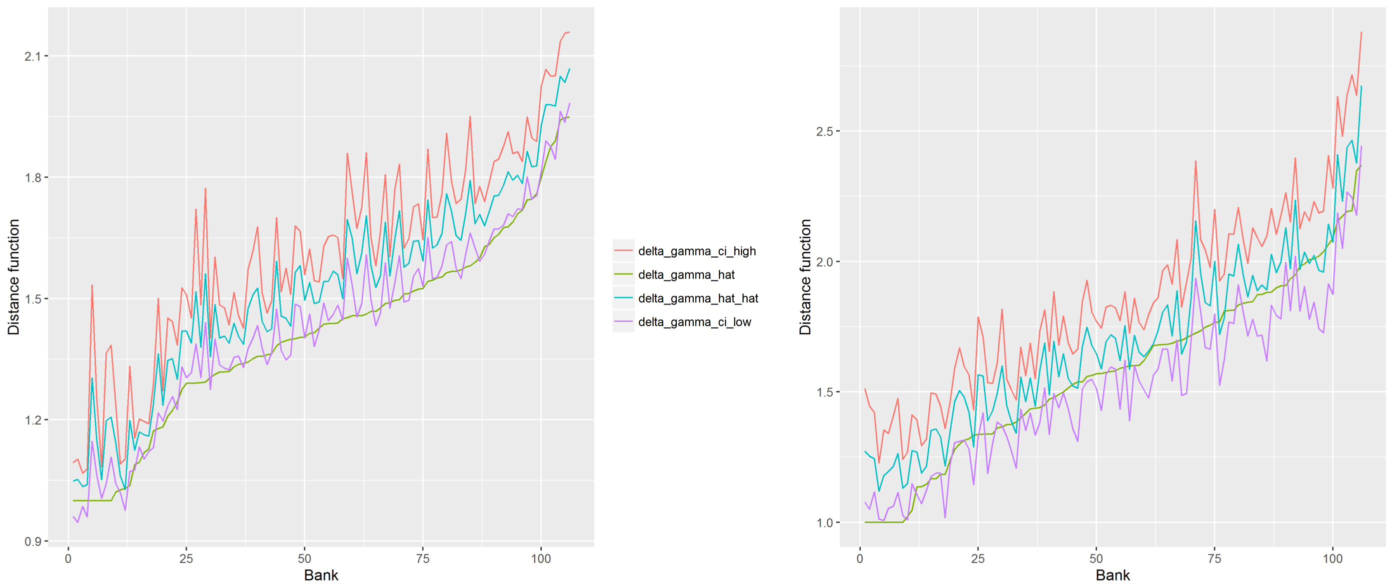

Figure 8: **"Naive" and bias-corrected distance function scores for models with asset approach (left) and intermediation approach (right) in presence of environmental variables**

6 Conclusion

The paper shows that a direct modification of Simar and Wilson (1998) and Simar and Wilson (2007) methodology is inconsistent for correcting the bias of Fare et al. (1985) cost efficiency scores and proposes an alternative bootstrap algorithm for robust estimation. To approximate the bias of "naive" cost efficiency score, 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 simulation analyses for multi-input multi-output Cobb-Douglas production function with correlated outputs, and correlated technical and cost efficiency, show consistency of the proposed algorithm in terms of coverage probability of Kneip et al. (2008) confidence intervals for true cost efficiency. Consistency generally holds even for small samples. An application of the algorithm to real data of 106 Japanese banks in the fiscal year 2013 demonstrates re-ranking of banks according to their bias-corrected cost efficiency scores, as well as shows heterogeneity of bias according to bank charter.

Appendix A 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 n-th input, used to produce m-th output ($x_n = \sum_{m=1}^{M} x_{nm}$, 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 (Shephard (1981); Resti (2000)):

$$x^*_{nm} = \frac{(y^*_m/A_m)^{1/\sum_{n=1}^{N}\alpha_{nm}}\alpha_{nm}}{\prod_{n=1}^{N}\alpha_{nm}{}^{\alpha_{nm}/\sum_{n=1}^{N}\alpha_{nm}}} \Bigg/ \frac{w_n}{\prod_{n=1}^{N} w_n{}^{\alpha_{nm}/\sum_{n=1}^{N}\alpha_{nm}}} = \frac{(y^*_m/A_m)^{1/\rho_m}\alpha_{nm}T_m}{w_n}, \tag{A.2}$$

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 did not change (Resti (2000)). Formally,

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

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

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:

$$\gamma = \frac{\mathbf{w}\mathbf{x}^{opt}}{\mathbf{w}\mathbf{x}} = \frac{\sum_{n=1}^{N} x_n^{opt} w_n}{\sum_{n=1}^{N} x_n w_n} \tag{A.6}$$

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

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

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

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

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

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 $y_m = y_m^* \theta^{\rho_m}$, we can rewrite

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

Then, cost efficiency γ is calculated as follows:

$$\gamma = \frac{\mathbf{w}\mathbf{x}^{opt}}{\mathbf{w}\mathbf{x}} = \frac{\sum_{n=1}^{N} \sum_{m=1}^{M} (y_m^*/A_m)^{1/\rho_m} \theta \alpha_{nm} T_m}{\sum_{n=1}^{N-1} \sum_{m=1}^{M} (y_m^*/A_m)^{1/\rho_m} \alpha_{nm} T_m \eta_{nm} + \sum_{m=1}^{M} (y_m^*/A_m)^{1/\rho_m} \alpha_{Nm} T_m \prod_{n=1}^{N-1} \eta_{nm}^{-\alpha_{nm}/\alpha_{Nm}}} \tag{A.13}$$

Our estimations with Japanese data and the results in the empirical literature (Liu et al. (2012); Wang (2003); Banker et al. (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:

$$\gamma = \frac{T\theta \sum_{n=1}^{N} \alpha_n \sum_{m=1}^{M} (y_m^*/A_m)^{1/\rho}}{T \sum_{n=1}^{N-1} \alpha_n \eta_n \sum_{m=1}^{M} (y_m^*/A_m)^{1/\rho} + T\alpha_N \prod_{n=1}^{N-1} \eta_n^{-\alpha_n/\alpha_N} \sum_{m=1}^{M} (y_m^*/A_m)^{1/\rho}} \tag{A.14}$$

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

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

References

Altunbas, Y., Liu, M.-H., Molyneux, P., and Seth, R. (2000). Efficiency and risk in Japanese banking. *Journal of Banking and Finance*, 24(10):1605–1628.

Aly, H., Grabowski, R., Pasurka, C., and Rangan, N. (1990). Technical, scale, and allocative efficiencies in U.S. banking: an empirical investigation. *Review of Economics and Statistics*, 72:211–217.

Badin, L. and Simar, L. (2003). Confidence intervals for DEA-type efficiency scores: how to avoid the computational burden of the bootstrap. Technical report 0322, IAP Statistics Network, http://sites.uclouvain.be/IAP-Stat-Phase-V-VI/PhaseV/publications_2003/TR/TR0322.pdf.

Banker, R., Gadh, V., and Gorr, W. (1993). A Monte Carlo comparison of two production frontier estimation methods: Corrected ordinary least squares and data envelopment analysis. *European Journal of Operational Research*, 67:332–343.

Barros, C. and Dieke, P. (2008). Measuring the economic efficiency of airports: A Simar-Wilson methodology analysis. *Transportation Research Part E*, 44:1039–1051.

Barros, C., Managi, S., and Matousek, R. (2012). The technical efficiency of the Japanese banks: Non-radial directional performance measurement with undesirable output. *Omega*, 40:1–8.

Besstremyannaya, G. (2013). The impact of Japanese hospital financing reform on hospital efficiency. *Japanese Economic Review*, 64(3):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*, 42:66–89.

Charnes, A., Cooper, W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2:429–444.

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

de Borger, B., Kerstens, K., and Staat, M. (2008). Transit costs and cost efficiency: Bootstrapping non-parametric frontiers. *Research in Transportation Economics*, 23:53–64.

Drake, L. and Hall, M. (2003). Efficiency in Japanese banking: an empirical analysis. *Journal of Banking and Finance*, 27:891–917.

Drake, L., Hall, M., and 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*, 19:1–15.

Fare, R., Grosskopf, S., and Lovell, C. (1985). *The Measurement of Efficiency in Production.* Kluwer Academic Publishers, Boston.

Farrell, M. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society, Series A (General)*, 3:253–281.

Fethi, M. and Pasiouras, F. (2010). Assessing bank efficiency and performance with opeartional research and artificial intelligence techniques: A survey. *European Journal of Operational Research*, 204:189–198.

Fukuyama, H. (1993). Technical and scale efficiency of Japanese commercial banks: a non-parametric approach. *Applied Economics*, 25:1101–1112.

Fukuyama, H. (1995). Measuring efficiency and pruductivity growth in Japanese banking: a nonparametric frontier approach. *Applied Financial Economics*, 5:95–107.

Fukuyama, H. and Weber, W. (2002). Estimating output allocative efficiency and productivity change: application to Japanese banks. *European Journal of Operation Research*, 137:177–190.

Fukuyama, H. and Weber, W. (2008). Japanese banking inefficiency and shadow pricing. *Mathematical and Computer Modelling*, 48:1854–1867.

Fukuyama, H. and Weber, W. (2010). A slacks-based inefficiency measure for a two-stage system with bad outputs. *Omega*, 38:398–409.

Giokas, D. (1991). Bank branch operating efficiency: A comparative application of DEA and the loglinear model. *Omega*, 19(6):549–557.

Glass, J., McKillop, D. G., and Morikawa, Y. (1998). Intermedication and value-added models for estimating cost economies in large Japanese banks 1977-93. *Applied Financial Economics*, 8:653–661.

Hori, K. and Yoshida, A. (1996). Nihon-no ginkougyou-no hiyou kouritsusei [Measuring cost efficiency in Japanese banking]. *Japanese Journal of Financial Economics*, 1:87–110.

Kasuya, M. (1986). Economies of scope: theory and application to banking. *BOJ Monetary and Economic Studies*, 4:59–104.

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*, 8:79–118.

Kittelsen, S. (1999). Monte Carlo simulations of DEA efficiency measures and hypothesis tests. Memorandum 09, Department of Economics, University of Oslo.

Kneip, A., Simar, L., and Wilson, P. (2008). Asymptotics and consistent bootstraps for DEA estimators in nonparametric frontier models. *Econometric Theory*, 24:1663–1697.

Kneip, A., Simar, L., and Wilson, P. (2011). A computationally efficient, consistent bootstrap for inference with non-parametric DEA estimators. *Computational Economics*, 38:483–515.

Kumbhakar, S. (2011). Estimation of Multiple Output Production Functions. 2012 spring seminar series April 17, Georgia Tech, http://www.econ.gatech.edu/files/seminars/kumbhakar11.pdf.

Linna, M., Hakkinen, U., Peltola, M., Magnussen, J., Anthun, K., Kittelsen, S., Roed, A., Olsen, K., Medin, E., and Rehnberg, C. (2010). Measuring cost efficiency in the Nordic Hospitals – a cross-sectional comparison of public hospitals in 2002. *Health Care Management Science*, 13:346–357.

Liu, J. and Tone, K. (2008). A multistage method to measure efficiency and its application to Japanese banking industry. *Socio-Economic Planning Sciences*, 42:75–91.

Liu, L., Ondrich, J., and Ruggiero, J. (2012). Estimating multiple-input–multiple-output production functions with an analysis of credit unions. *Applied Economics*, 44:1583–1589.

McKillop, D., Glass, J., and Morikawa, Y. (1996). The composite cost function and efficiency in giant Japanese banks. *Journal of Banking and Finance*, 20:1651–1671.

Medin, E., Anthum, K., Hakkinen, U., Kittelsen, S., Linna, M., Magnussen, J., Olsen, K., and Rehnberg, C. (2011). Cost efficiency of university hospitals in the Nordic countries: a cross-country analysis. *European Journal of Health Economics*, 12(6):509–519.

Miyakoshi, T. and Tsukuda, Y. (2004). Regional disparities in Japanese banking performance. *RURDS*, 16(1):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*, 121:559–578.

Shephard, R. W. (1981). Cost and production functions. In Beckmann, M. and Kunzi, H., editors, *Lecture Notes in Economics and Mathematical Systems*, Berlin Heidelberg New York. Springer-Verlag.

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

Simar, L. and Wilson, P. (1998). Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. *Management Science*, 44:49–61.

Simar, L. and Wilson, P. (2000a). A general methodology for bootstrapping in non-parametric frontier models. *Journal of Applied Statistics*, 27(6):779–802.

Simar, L. and Wilson, P. (2000b). Performance of the bootstrap for DEA estimators and iterating the principle. STAT Discussion Papers 0002.

Simar, L. and Wilson, P. (2002). Non-parametric tests of returns to scale. *European Journal of Operational Research*, 139(1):115–132.

Simar, L. and Wilson, P. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, 136:31–64.

Simar, L. and Wilson, P. (2008). Statistical inference in nonparametric frontier models: recent developments and perspectives. In O.Fried, H., Lovell, C., and Schmidt, S., editors, *The Measurement of Productive Efficiency and Productivity Growth*, pages 421–521. Oxford University Press.

Simar, L. and Wilson, P. (2011a). Inference by the m out of n bootstrap in nonparametric frontier models. *Journal of Productivity Analysis*, 36:33–53.

Simar, L. and Wilson, P. (2011b). Two-stage DEA: caveat emptor. *Journal of Productivity Analysis*, 36:205–218.

Simm, J. and Besstremyannaya, G. (2016). Robust Data Envelopment Analysis (DEA) for R, https://cran.r-project.org/web/packages/rdea/index.html.

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*, 479:598–602.

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

Tone, K. (2002). A strange case of the cost and allocative efficiencies in DEA. *Journal of the Operational Research Society*, 53(11):1225–1231.

Tortosa-Austina, E. (2002). Bank cost efficiency and output specification. *Journal of Productivity Analysis*, 18:199–222.

Wang, J. (2003). Productivity and economies of scale in the production of bank service value added. Working Paper September, Research Department, Federal Reserve Bank of Boston, http://www.bostonfed.org/economic/wp/wp2003/wp037.pdf.