FDH Efficiency Scores from a Stochastic Point of View

B. U. Park [*, ‡]
L. Simar [*,§]
Ch. Weiner [*]

August 1997

Abstract

The Free Disposal Hull (FDH) is a nonparametric estimator for the production set. In Productivity Analysis one derives the production frontier and efficiency scores from the FDH. In the literature the method is considered to be deterministic. However, assuming that individuals are drawn independently from a distribution, where the support is the true production set, FDH efficiency scores are random variables. The paper investigates its stochastic properties.

[*]Institute de statistique, Université catholique de Louvain, Louvain-la-Neuve, Belgium, supported by "Projet d'Actions de Recherche Concertées" (No. 93/98 - 164)

[‡]Department of Statistics, Seoul National University, Seoul, Korea; supported by the Nondirected Research Fund, Korea Research Foundation, 1996.

[§]CORE, Université catholique de Louvain, Louvain-la-Neuve, Belgium

1 Introduction

In Productivity and Efficiency Analysis, we are interested in the *production* or *technology set* Ψ, i. e. a set of technically possible pairs of input $x \in I\!R_+^p$ and output $y \in I\!R_+^q$. We deal with this set, because we are searching the *production frontier*, i. e. the set of the most efficient production processes, and the *efficiency* of an individual, i. e. the distance of its in- and output to the best alternative.

If we have no or only vague information about technical possibilities and restrictions, we can still estimate the production set from the observed population. There are two nonparametric methods to estimate Ψ: The *Data Envelopment Analysis* (DEA) and the *Free Disposal Hull* (FDH). Both estimators cover the data with the smallest set, that has some typical properties of a production set. As the FDH also envelopes the data it is sometimes considered to be a special case of data envelopment. In applications to real data the question whether to choose DEA or FDH depends on the properties which we can assume. Comparing with the FDH, the DEA requires stronger assumptions, hence the FDH is a more general estimator. The method is due to Farrell [7] (1957) and Charnes, Cooper and Rhodes [3] (1978); both dealed with a DEA. The FDH was suggested in 1984 by Deprins, Simar and Tulkens for the analysis of public enterprises.

Estimators for the production set also induce estimators for the frontier and the efficiencies, if we use the estimated production set as empirical standard. This method was also proposed by Farrell. He used a *radial distance* from the output to the frontier as efficiency measure. This radial distance either of the output or the input is nowadays the most popular efficiency measure in applications.

A lot of econometricians consider that DEA and FDH are deterministic methods. From this point of view, the DEA or FDH of the whole population is the production set and nothing more. But we can also consider a stochastic model, where individuals are drawn according to a probability distribution with unknown support Ψ. Then we have to take into account two effects: DEA and FDH are too small, and there is random variation in the estimates.

There are already some results known about the statistical properties of data envelopment: Afriat [1] (1972) and Banker [2] (1993) consider multivariate input $p > 1$ and univariate output $q = 1$. In this setup, the production frontier is the graph of a production function $g : I\!R^p \to I\!R$ and the inefficiency is a half sided error term. Afriat deals with some parametric models for the inefficiency, Banker considers a monotonically decreasing density for the inefficiency. Korostelev, Simar and Tsybakov [9] and [10] (1995) analyze the FDH and DEA estimators for the production set for multivariate input and univariate output, too. Simar and Wilson [12] propose a bootstrap algorithm to derive numerical expressions for bias and confidence bounds for the efficiency scores. The algorithm can be applied for multivariate in- and output. Kneip, Park and Simar (1996) [8] analyze DEA efficiency scores of a production plan (x_0, y_0); they derive the rate of convergence in the case of multivariate in- and output. Gijbels, Mammen, Park and Simar (1996) [6] derive the asymptotic distribution of the DEA estimator for the production function for univariate in- and output.

The present paper complements the results of the latter two publications. We show the rate of convergence and the asymptotic distribution of the FDH estimator for the efficiency score of a pair (x_0, y_0). Using this asymptotic distribution we propose a bias corrected estimator and asymptotic confidence bounds. Rate of convergence, bias and variance of FDH estimates, which we find, are worse than the errors of DEA estimates. This is an obvious result, because the FDH is more general than the DEA. But this is not the important point of this work. Moreover we are able to derive the asymptotic distribution in the case of multivariate in- and output, a results that is still missing for the DEA.

The assumptions we use are some weak regularity conditions on the production set and the distribution of the individuals. They are common for nonparametric estimation and less restrictive than any parametric model. We make no assumptions on the dependence or independence of the different inputs and outputs, but we assume that individuals are independently drawn from a population.

We do not consider the problem, where we have an additional noise on the observations; there we have an identification problem. The model is identified either by strong parametric assumptions or repeated observations for every individual. Such a work stays in its own right.

The paper is organized as follows: In section 2 we start with a brief review of frontier analysis. Section 3 provides three results: we show that estimates for the efficiency scores follow a Weibull distribution with parameters $\mu_{NW} n^{\frac{1}{p+q}}$ and $p+q$, we propose an estimator for the parameter μ_{NW}, and we construct a bias corrected estimator for ϑ_0 and asymptotic confidence bounds. Exemplarily there are some simulations and an application to real data in section 4. Finally there is an appendix with lemmas and proofs.

2 Frontier Analysis

2.1 Production set

The *production* or *technology set* Ψ is the set of all feasible *production plans*, i. e. pairs of in- and output $(x, y) \in I\!R_+^{p+q}$. For $p > 1$ or $q > 1$ we denote $x = (x^1, \dots x^p)$, respectively $y = (y^1, \dots, y^q)$, where x^k and y^k are the different input and output variables. Comparing two pairs we say the one is at least as efficient as the other, if it produces at least as much output with the same or less input. We assume that the production set is *free disposal*, i. e. if a pair (x, y) lies in Ψ, then all less efficient pairs (x', y') are also contained in Ψ. For technical reasons we also assume that Ψ is compact.

One can further assume that the production set is convex. That means that every convex combination of feasible production plans is also feasible. This assumption is widely used, but it is not generally valid. The production technology might admit *increasing returns to scale*, i. e. the output increases faster than the inputs, or there might be lumpy goods, i. e. fractional values of inputs or outputs do not exists.

Let $y^{(k)} := (y^1, \dots, y^{k-1}, y^{k+1}, \dots, y^q)$ denote the vector y without the $k\,th$ component and $y^{(k)}(\eta) := (y^1, \dots, y^{k-1}, \eta, y^{k+1}, \dots, y^q)$ the vector where we insert an η as $k\,th$ component into the vector $y^{(k)}$. For every $k : 1 \leq k \leq q$, we can define a function

$$g_\Psi^k(x, y^{(k)}) \;:=\; \max\{\,\eta | (x, y^{(k)}(\eta)) \in \Psi\}.$$

As Ψ is a free disposal set, these functions are monotonically increasing in x and monotonically decreasing in $y^{(k)}$. On the other hand: for an arbitrary function $\tilde{g}$, that is monotonically increasing in x and monotonically decreasing in $y^{(k)}$, we can define a set

$$\Psi^k(\tilde{g}) \;:=\; \{\,(x, y^{(k)}(\eta)) | \eta \leq \tilde{g}(x, y^{(k)})\,\}.$$

One can easily check that $\Psi^k(g_\Psi^k) = \Psi$ (lemma A.1, see Appendix A).

Obviously we can do the same things with functions $h_\Psi^k(x^{(k)}, y) := \min\{\xi | (x^{(k)}(\xi), y) \in \Psi\}$. Thus a production set Ψ is uniquely determined by either of the functions g_Ψ^k or h_Ψ^k. If we deal with univariate output, then there is only one function $g_\Psi : I\!R^p \to I\!R$. In this case, g_Ψ is called the *production function*.

For the rest of the paper we simplify the notation: we only consider the function g_Ψ^q, which we denote by g. Also, we fix a particular pair (x_0, y_0). The efficiency measure of the pair (x_0, y_0) and its estimator are defined in the next two subsections.

2.2 Efficiency

The *efficiency*, or *efficiency score* is a measure that indicates the distance to the best alternative. The most popular measure is the *Farrell efficiency measure*, which was introduced by Farrell [7] in 1957. It is a radial distance, i. e. for an observed input we compare the observed output with all outputs on the ray $\{\vartheta y_0 | (x_0, \vartheta y_0) \in \Psi\}$. The largest scale factor ϑ indicates the efficiency.

$$\vartheta_\Psi^{\mathrm{out}}(x_0, y_0) := \max\{\, \vartheta | (x_0, \vartheta y_0) \in \Psi \,\} \tag{1}$$

We may also interpret $1/\vartheta$ as the fraction of the efficient alternative, which the production plan (x_0, y_0) achieves.

As this measure compares possible output, it is called the *(Farrell) output efficiency measure.* In the same way we can define an input efficiency measure $\vartheta_\Psi^{\mathrm{in}}(x_0, y_0) := \min\{\vartheta | (\vartheta x_0, y_0) \in \Psi\}$. The discussion of the properties of input efficiency is absolutely equivalent to the discussion of output efficiency, therefore we restrict the discussion to output efficiency. For more details we reffer to Färe, Grosskopf and Lovell [5].

We denote the efficient output or efficient alternative for (x_0, y_0) by

$$y_0^\partial := \vartheta_\Psi^{\mathrm{out}}(x_0, y_0) y_0$$

The set of all y_0^∂, $(x_0, y_0) \in \Psi$, forms the *(Farrell) production frontier.* Since we can compute the efficiency for all production plans $(x_0, y_0) \in \Psi$, the frontier is uniquely determined.

For $q = 1$ we have: $\vartheta_\Psi^{\mathrm{out}}(x, y) = g(x)/y$ and the production frontier is the graph of g.

Again we simplify the notation and denote $\vartheta_0 := \vartheta_\Psi^{\mathrm{out}}(x_0, y_0)$.

2.3 Free Disposal Hull (FDH)

Unfortunately the production set Ψ is unknown. However, we have independent observations from which we can estimate Ψ. A well known estimator is the *Free Disposal Hull* (FDH)

Definition 2.1 *Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ denote the observed production plans. The free disposal hull (FDH) is the smallest free disposal set containing all observations. It is given by*

$$\Psi_{FDH} := \{\, (x, y) \,|\, \exists i : \, x \geq X_i \textit{ and } y \leq Y_i \,\}.$$

For comparison, the DEA is the smallest free disposal and convex set that covers all data, i. e. the convex hull of the FDH.

Using the FDH, we derive naive estimators for the efficiency, $\widehat{\vartheta_0} := \max\{\vartheta | (x_0, \vartheta y_0) \in \Psi_{FDH}\}$, and the efficient output, $\widehat{y_0^\partial} := \widehat{\vartheta_0} y_0$. Consider that $\vartheta y_0 \leq Y_i$ if and only if $\vartheta \leq \min_{1\leq k\leq q} Y_i^k / y_0^k$, hence

$$\widehat{\vartheta_0} := \max_{i:X_i \leq x_0} \min_{1\leq k \leq q} Y_i^k / y_0^k \tag{2}$$

3 The main results

3.1 Asymptotic distribution of $\widehat{\vartheta_0}$

For the probabilistic setup we assume

Assumption A I *The observations (X_i, Y_i) are iid*

Assumption A II *The common distribution of (X_i, Y_i) has a density f. At the frontier this density is positive, i. e. $f_0 := f(x_0, y_0^\partial) > 0$, and sequentially Lipschitz continuous, i. e. for all sequences (x_n, y_n) in Ψ converging to (x_0, y_0^∂) holds: $|f(x_n, y_n) - f(x_0, y_0^\partial)| \leq C_1 \|(x_n, y_n) - (x_0, y_0^\partial)\|$ for some positive C_1*

Moreover we need a regularity condition on the production set:

Assumption A III *At the frontier, $g(\cdot,\cdot)$ is positive, i. e. $g(x_0, y_0^{\partial(q)}) > 0$, continuously differentiable, the first derivative is Lipschitz continuous, i. e. for all $(x, y) : |g(x, y^{(q)}) - g(x_0, y_0^{\partial(q)}) - \nabla g(x_0, y_0^{\partial(q)})^t((x, y^{(q)}) - (x_0, y_0^{(q)}))| \leq C_2 \|(x, y^{(q)}) - (x_0, y_0^{(q)})\|^2$, and for $k = 1, \ldots, p$ and $l = 1, \ldots, q-1$:*

$$\frac{\partial}{\partial x^k} g(x, y^{(q)})|_{(x,y^{(q)})=(x_0,y_0^{\partial(q)})} > 0 \ \ and \ \ \frac{\partial}{\partial y^l} g(x, y^{(q)})|_{(x,y^{(q)})=(x_0,y_0^{\partial(q)})} < 0.$$

In order to derive the asymptotic distribution of $\widehat{\vartheta_0}$ we approximate the cumulative distribution function. Consider an arbitrary $z > 0$

$$P[\vartheta_0 - \widehat{\vartheta_0} \leq z] \ = \ 1 - P[\max_{i:X_i \leq x_0} \min_{1\leq k\leq q} Y_i^k / y_0^k < \vartheta_0 - z]$$

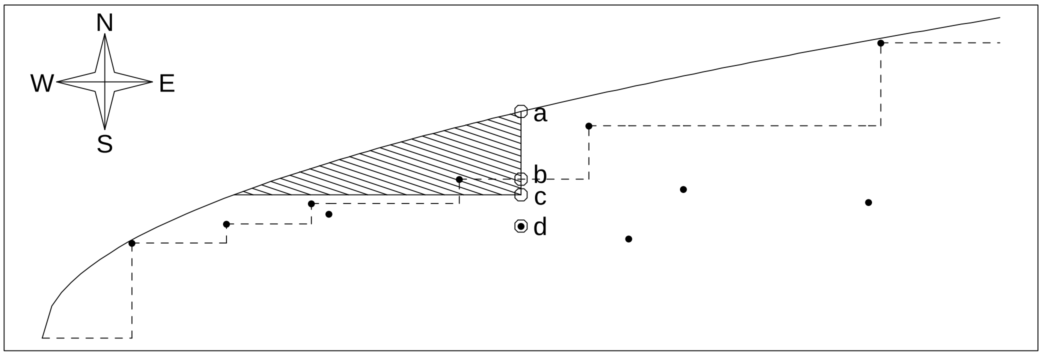

Figure 1: *Illustration for a FDH with $p = 1$ inputs and $q = 1$ outputs. The production set Ψ is the set under the solid line; this solid line is also the graph of $g(\cdot)$. Dots represent observed pairs (X_i, Y_i) and the dashed line is the frontier of the FDH; a, b, c and d label the points $(x_0, g(x_0))$, $(x_0, \widehat{g}(x_0))$, $(x_0, g(x_0) - zy_0)$, and (x_0, y_0). Consider estimating ϑ_0 : $\widehat{\vartheta_0}$ is bigger than $\vartheta_0 - z$ if and only if $\widehat{g}(x_0)$ is bigger than $g(x_0) - zy_0$ and this is equivalent to the case that there is an observation in "North West of $(x, g(x) - zy_0)$" (shaded area)*

The maximum over all $i : X_i \leq x_0$ of rv's $\tilde{Y}_i$ is smaller than a value z if we have for every i either[1] $X_i \not\leq x_0$ or $\tilde{Y}_i < z$. Since we assumed independent observations, we get

$$\begin{aligned} P[\vartheta_0 - \widehat{\vartheta_0} \leq z] &= 1 - (P[X_i \not\leq x_0 \textit{ or } \min_{1 \leq k \leq q} Y_i^k / y_0^k < \vartheta_0 - z])^n \\ &= 1 - (1 - P[X_i \leq x_0 \textit{ and } Y_i \geq y_0^\partial - zy_0])^n \qquad (3) \end{aligned}$$

As we already see in equation (3), $\widehat{\vartheta_0}$ is a consistent estimator for ϑ_0: $P[X_i \leq x_0 \textit{ and } Y_i \geq y_0^\partial - zy_0]$ is strictly positive if and only if z is strictly positive. Therefore the right side converges to one for strictly positive z and it converges to zero for $z = 0$.

Figure 1 shows an illustration for this transformation for the case of $p = q = 1$. Because of this geometric representation, we denote

$$NW(x_0, y_0^\partial - zy_0) := \{X_i \leq x_0 \textit{ and } Y_i \geq y_0^\partial - zy_0\} \cap \Psi$$

and call the set "North West of $(x_0, y_0^\partial - zy_0)$". Furthermore we denote

$$p_{NW}(z) := P[X_i \leq x_0 \textit{ and } Y_i \geq y_0^\partial - zy_0]$$

[1]By $X_i \not\leq x_0$ we mean X_i is not $\leq x_0$. Note that this does not mean $X_i > x_0$ because of the multivariate setting

These probabilities contain all information about the distribution of $\widehat{\vartheta_0}$; hence the crucial point is to approximate these sets as n tends to infinity.

In terms of productivity analysis, $NW(x_0, y_0) = NW(x_0, y_0^\partial - (1-\widehat{\vartheta_0})y_0)$ is the set of all production plans which are more efficient than (x_0, y_0), and $p_{NW}(1-\widehat{\vartheta_0})$ is the expected proportion of more efficient production plans. This number provides a measurement for the efficiency: the smaller $p_{NW}(1-\widehat{\vartheta_0})$ the more efficient the production plan (x_0, y_0). Hence $p_{NW}(\cdot)$ is interesting, also for non statistical purposes.

Asymptotically the distance $\vartheta_0 - \widehat{\vartheta_0}$ goes to zero, therefore we need the density f and the curve[2] g in a region around (x_0, y_0^∂), only. Let $\zeta = o(1)$ denote a sequence that goes to zero. Locally g is linear and f is constant. Hence $NW(x_0, y_0^\partial - \zeta y_0)$ is asymptotically a $p+q$ dimensional simplex where the angles depend on the partial derivatives of g in $(x_0, y_0^{\partial(q)})$, and $p_{NW}(\zeta)$ is approximately proportional to ζ^{p+q}. In lemma A.3 ("NW Lemma"; see Appendix A) we show

$$p_{NW}(\zeta) \; = \; (\mu_{NW}\zeta)^{p+q} + O(\zeta^{p+q+1}) \tag{4}$$

An algebraic expression for μ_{NW}, that depends on the distribution of the population is given in the appendix.

As $(1-p/n)^n$ converges to $\exp(-p)$, we get the asymptotic distribution function:

Theorem 3.1 *Assume A I, A II and A III. Then for all $z > 0$ holds*

$$P[n^{1/(p+q)}(\vartheta_0 - \widehat{\vartheta_0}) \le z] \; = \; 1 - \exp\Big[-(\mu_{NW}z)^{p+q}\Big] + o(1). \tag{5}$$

Corollary 3.2 *(i) $\widehat{\vartheta_0}$ converges to ϑ_0 with $O_P(n^{-1/(p+q)})$ and there is no $\alpha > 1/(p+q)$ such that $\vartheta_0 - \widehat{\vartheta_0} = O_P(n^{-\alpha})$*
(ii) Asymptotically $(\vartheta_0 - \widehat{\vartheta_0})$ follows a Weibull distribution[3] with parameters $\mu_{NW}n^{1/(p+q)}$ and $p+q$:

$$\vartheta_0 - \widehat{\vartheta_0} \; \sim \; AW(\mu_{NW}n^{1/(p+q)}, p+q)$$

Remark *Kneip et al. [8] found under similar assumptions that the DEA estimator has a faster rate of convergence, i. e. $\vartheta_0 - \widehat{\vartheta_0}^{DEA} = O_P(n^{-2/(p+q+1)})$. This is the price we have to pay for the more general setup, i. e. that we allow non-convexity of Ψ.*

[2] g is the function which we have defined in section 2.1

[3] Def: $X \sim W(\lambda, c) \Leftrightarrow X^{1/c} \sim Exp(\lambda)$

$p+q$	mean / σ	$p+q$	mean / σ
1	1.00	6	5.16
2	1.91	7	5.95
3	2.75	8	6.74
4	3.56	9	7.53
5	4.37	10	8.31

Table 1: *The ratio of mean and standard deviation of a $W(\lambda, p+q)$ distribution. As the dimension increases, the mean gets an increasing dominance*

The moments of a Weibull distributed random variable $X \sim W(1, p+q)$ are

$$E[X^r] \;=\; \Gamma\Big((p+q+r)/(p+q)\Big)$$

where $\Gamma(\cdot)$ denotes the Gamma function. We can use these moments as approximations for the moments of $(\vartheta_0 - \widehat{\vartheta_0})$. However, the convergence in distribution of $n^{1/(p+q)}(\vartheta_0 - \widehat{\vartheta_0})$ (theorem 3.1) does not trivially imply the convergence of the moments. Therefore we proof the following theorem in the appendix:

Theorem 3.3 *Assume A I, A II and A III. It holds*

$$E[(\vartheta_0 - \widehat{\vartheta_0})^r] \;=\; c_r \mu_{NW}^{-r} n^{-r/(p+q)} + o(n^{-r/(p+q)})$$

where $c_r = \Gamma\Big((p+q+r)/(p+q)\Big)$

At this point we can already give some comments on these results:

1) The speed of convergence goes down as the dimension of the data, $p+q$, increases. Thus, estimating efficiency scores with a FDH, we need a sample size which exponentially depends on $p+q$. This reflects the well known *curse of dimensionality* in nonparametric curve estimation.

2) A $W(\mu_{NW} n^{1/(p+q)}, p+q)$ distribution has mean $c_1 \mu_{NW}^{-1} n^{-1/(p+q)}$ and standard deviation $\sqrt{c_2 - c_1^2}\, \mu_{NW}^{-1} n^{-1(p+q)}$. The ratio of mean and standard deviation does not depend on μ_{NW} and n, in table 1 we give numerical expressions for this ratio for dimensions $p+q = 1, \ldots 10$. We see that the bias is at least as large as the standard deviation and that this ratio increases as the dimension $p+q$ of the data increases. This is a second deterioration of the estimator besides the curse of dimensionality. However, we can estimate this bias and correct for it. As table 1 indicates, it is worth to do it!

3.2 Estimating the parameter μ_{NW}

The asymptotic distribution has the unknown parameter μ_{NW}, which we can estimate from the data. First we estimate $p_{NW}(\zeta)$ by the proportion of observations North West of $(x_0, \widehat{y_0^\partial} - \zeta y_0)$:

$$\widehat{p_{NW}}(\zeta) := \frac{\sharp\{\,(X_i, Y_i) \,:\, (X_i, Y_i) \in NW(x_0, \widehat{y_0^\partial} - \zeta y_0)\,\}}{n}, \tag{6}$$

Actually we are interested in the observations North West of $(x_0, y_0^\partial - \zeta y_0)$, but y_0^∂ is unknown. Therefore we replace it by the FDH estimate $\widehat{y_0^\partial}$, which is a consistent estimate. However, we have to take into account the error of $\widehat{\vartheta_0}$. Consider $\widehat{y_0^\partial} - \zeta y_0 = y_0^\partial - \zeta y_0 + O_P(n^{-1/(p+q)})$. We can ignore the O_P term, if it is much smaller than $y_0^\partial - \zeta y_0$; asymptotically this corresponds to a ζ that goes slower to zero than $n^{-1/(p+q)}$. Mean and variance of $\widehat{p_{NW}}(\zeta)$ are calculated in lemma A.4; see appendix A.

From the NW lemma we now derive a naive estimator for μ_{NW}

$$\widehat{\mu_{NW}} := \zeta^{-1}\widehat{p_{NW}}(\zeta)^{1/(p+q)}. \tag{7}$$

This estimator has the following bias and variance:

Theorem 3.4 *Assume A I, A II and A III and consider $\zeta \to 0$, such that $n^{\frac{1}{p+q}}\zeta \to \infty$. Bias and variance of $\widehat{\mu_{NW}}$ are*

$$\begin{aligned} bias[\widehat{\mu_{NW}}] &= c_1\zeta^{-1}(n-1)^{-1/(p+q)} + o(\zeta^{-1}n^{-1/(p+q)}) + O(\zeta) \\ var[\widehat{\mu_{NW}}] &= O(\zeta^{-2}n^{-2/(p+q)}) + O(\zeta) \end{aligned}$$

where $c_1 = \Gamma\Big((p+q+1)/(p+q)\Big)$

Remark *Taking $\zeta = O(n^{-2/(3(p+q))})$, we get $MSE(\widehat{\mu_{NW}}) = O(n^{-2/(3(p+q))})$. One can also show that this rate cannot be improved by choosing a faster or slower ζ.*

Yet in a finite sample for small ζ, $\widehat{\mu_{NW}}$ has bad properties. Consider that there might be no observation in $NW(x_0, y_0^\partial - \zeta y_0)$, but there is at least one observation in $NW(x_0, \widehat{y_0^\partial} - \zeta y_0)$, particularly the observation which induces the estimated frontier. Thus the estimator has a positive bias for too small ζ.

This bias is strongly related with the bias of $\widehat{\vartheta_0}$ and corresponds to the $c_1\zeta^{-1}(n-1)^{-1/(p+q)}$ term in theorem 3.4. Correcting this bias does not necessarily yield a faster estimator for $\widehat{\mu_{NW}}$, but it improves the quality of the estimates for small samples.

Finally we propose the following estimator for μ_{NW}: Let $\zeta = Cn^{-2/(3(p+q))}$ (cf. the remark for theorem 3.4 above)

$$\widehat{\widehat{\mu_{NW}}} := \zeta^{-1}\widehat{p_{NW}}(\zeta)^{1/(p+q)} - c_1\zeta^{-1}(n-1)^{-1/(p+q)} \tag{8}$$

ζ is a *bandwidth* parameter. $\widehat{\widehat{\mu_{NW}}}$ is consistent with an optimal rate of convergence for every fixed C. However, in a finite sample we are interested in a constant C that minimizes the error, or which makes it sufficiently small. This question arrives in many nonparametric curve estimation problems and it is quite delicate. We explain how one can choose this parameter in the practical part below (section 4)

3.3 A bias corrected estimator for ϑ_0 and asymptotic confidence bounds

A bias corrected estimator for ϑ_0 and asymptotic confidence bounds are straight forward consequences of the preceding two subsections: Since $\widehat{\widehat{\mu_{NW}}}$ is a consistent estimator for μ_{NW}, and since $\widehat{\vartheta_0} - c_1\mu_{NW}^{-1}n^{-1/(p+q)}$ converges with $o_P(n^{-1/(p+q)})$ to ϑ_0, we get

$$\widehat{\widehat{\vartheta_0}} := \widehat{\vartheta_0} - c_1\widehat{\widehat{\mu_{NW}}}^{-1}n^{-1/(p+q)} \tag{9}$$

For the asymptotic confidence bounds, let z_α denote the quantiles of a $W(1, p+q)$ distribution, i. e. $z_\alpha = -(\log(1-\alpha))^{1/(p+q)}$. Then we have

$$P[\widehat{\vartheta_0} \leq \vartheta_0 \leq \widehat{\vartheta_0} + \mu_{NW}^{-1}n^{-1/(p+q)}z_{1-\alpha}] \approx \alpha$$

and again as $\widehat{\widehat{\mu_{NW}}}$ is a consistent estimator, we may replace μ_{NW} by $\widehat{\widehat{\mu_{NW}}}$. Therefore the set

$$[\widehat{\vartheta_0}, \widehat{\vartheta_0} + \widehat{\widehat{\mu_{NW}}}^{-1}n^{-1/(p+q)}z_{1-\alpha}]$$

provides asymptotic confidence bounds for ϑ_0. As a Weibull distribution is bounded below by zero, we have chosen one sided confidence bounds.

4 Finite sample performance

Here we illustrate the results of the preceding section with simulations, and apply it to a real data set. We also give a data driven procedure for finding the parameter $\zeta = C\,n^{-2/(3(p+q))}$.

4.1 Simulation

We simulate with two setups:

simulation I *$p = q = 1$, $g_1(x) = 3(x-1.5)^3 + 0.25x + 1.125$, $X_i \sim U([1,2])$ and $Y_i = \exp(Z_i)g_1(X_i)$, where $Z_i \sim Exp(3)$; $\exp(Z_i)$ represents the inefficiency. All random numbers X_i and Z_i are drawn independently. We evaluate ϑ_0 for $(x_0, y_0) = (1.25, 1)$, $(1.5, 1)$ and $(1.75, 1)$.*

simulation II *$p = q = 2$, $g_2(x^1, x^2, y^1) = 2^{-1.4}(x^1)^{1.8}(x^2)^{0.6}(y^1)^{-1}$, $X_i^k \sim U([1,2])$, and $Y_i = \exp(Z_i)\vartheta(X_i, \tilde{Y}_i)\tilde{Y}_i$, where $\tilde{Y}_i^1 \sim U([0.2, 5])$, $\tilde{Y}_i^2 := \frac{1}{\tilde{Y}_i^1}$ and $Z_i \sim Exp(3)$. The points $(X_i^1, X_i^2, \vartheta(X_i, \tilde{Y}_i)\tilde{Y}_i^1, \vartheta(X_i, \tilde{Y}_i)\tilde{Y}_i^2)$ are uniformly distributed on the frontier and $\exp(Z_i)$ represents the inefficiency. All random numbers X_i^k, $\tilde{Y}_i^1$ and Z_i are drawn independently. We evaluate ϑ_0 for $(x_0, y_0) = (1.5, 1.5, 0.5, 0.5)$.*

Simulation I is an example for a non-convex production set. In simulation II we have chosen a generalized Cobb Douglas function with increasing returns to scale; it is an easy example for a non-convex multivariate production set. $y^2 = 2^{-1.4}(x^1)^{1.8}(x^2)^{0.6}(y^1)^{-1}$ is equivalent to $y^1y^2 = 2^{-0.7}(x^1)^{0.9}(x^2)^{0.3}$. On the right side the exponents have the typical relation of the exponents for input factors capital and labor. On the left side we have modeled two outputs of similar importance.

4.1.1 Asymptotic distribution versus empirical distribution

First we compare the Weibull distribution with parameters $\mu_{NW}n^{1/(p+q)}$ and $p+q$, which we derive from the simulation setup, with the empirical distribution of a FDH efficiency score derived from 500 simulations. In table 2 we compare the mean and the 5 % quantile for 5 different sample sizes, $n = 25$, 50, 100, 250, 500. For a sample size of 100 observations we show in figure 2 the Weibull density and a kernel estimate (Epanechnikov kernel and bandwidth according to Silverman's rule of thumb) of the empirical distibution.

In simulation I at $x_0 = 1.25$ and $x_0 = 1.5$, the asymptotic distribution overestimates mean and 5 % quantile, at $x_0 = 1.75$ it underestimates them. In the density plots we see, that the Weibull density is slightly too peaked for $x_0 = 1.25$ and $x_0 = 1.5$, whereas it is slightly too flat at $x_0 = 1.75$. We can explain these deviations: We approximate the region $NW(x_0, y_0^\partial - \zeta y_0)$ with a simplex. However, g_1 is not linear. In $x_0 = 1.25$ and $x_0 = 1.5$ it is concave. Here the simplex is smaller than $NW(x_0, y_0^\partial - \zeta y_0)$, hence μ_{NW} is too small and mean and 5% quantile are too large. At $x_0 = 1.75$, g_1 is convex

(x_0, y_0)	n	sim. mean	A mean	sim. 5% qtl	A 5% qtl
simulation I					
(1.25,1)	25	1.205	1.237	1.013	1.090
(1.25,1)	50	1.261	1.282	1.123	1.178
(1.25,1)	100	1.306	1.314	1.212	1.240
(1.25,1)	250	1.338	1.342	1.284	1.296
(1.25,1)	500	1.355	1.356	1.323	1.323
(1.5,1)	25	1.381	1.411	1.241	1.327
(1.5,1)	50	1.428	1.437	1.349	1.378
(1.5,1)	100	1.450	1.456	1.394	1.413
(1.5,1)	250	1.470	1.472	1.441	1.445
(1.5,1)	500	1.479	1.480	1.459	1.461
(1.75,1)	25	1.483	1.444	1.385	1.286
(1.75,1)	50	1.511	1.492	1.446	1.381
(1.75,1)	100	1.536	1.527	1.477	1.448
(1.75,1)	250	1.562	1.557	1.520	1.507
(1.75,1)	500	1.574	1.572	1.543	1.537
simulation II					
(1.5,1.5,0.5,0.5)	25	1.251	1.270	0.904	0.938
(1.5,1.5,0.5,0.5)	50	1.391	1.386	1.089	1.108
(1.5,1.5,0.5,0.5)	100	1.498	1.484	1.279	1.250
(1.5,1.5,0.5,0.5)	250	1.591	1.590	1.414	1.404
(1.5,1.5,0.5,0.5)	500	1.659	1.656	1.507	1.499

Table 2: *Distribution of FDH efficiency scores: mean and 5 % quantile of the empirical distribution (500 iterations) compared with mean and 5 % quantile of the approximating Weibull distribution. The Weibull distribution is a good approximation, also for small sample sizes.*

and, obviously, we see the opposite effects.

In simulation II, g_2 is neither convex nor concave in $(x_0, y_0^{\partial(2)}) = (1.5, 1.5, \theta_0\, 0.5)$. For small sample size, the Weibull distribution overestimates mean and 5 % quantile, for large sample size it underestimates it.

4.1.2 Estimating μ_{NW}

The second question which we investigate is: how good is our estimator for μ_{NW} and how to choose the $\zeta = C\, n^{-2/(3(p+q))}$. In figure 3 we see estimates with $\widehat{\mu_{NW}}$ (dotted line) and $\widehat{\widehat{\mu_{NW}}}$ (solid line) for different $\zeta = C\, n^{-2/(3(p+q))}$, $C \in (0, 3]$.

At the left side, for $C = 0.1$, $\widehat{\mu_{NW}}$ and $\widehat{\widehat{\mu_{NW}}}$ are very different. This shows the dominance of the "small sample bias of $\widehat{\mu_{NW}}$", i. e. $c_1 \zeta^{-1}(n-1)^{-1/(p+q)}$ (cf. theorem 3.4). At the right side for $C = 3$, the estimates are very close;

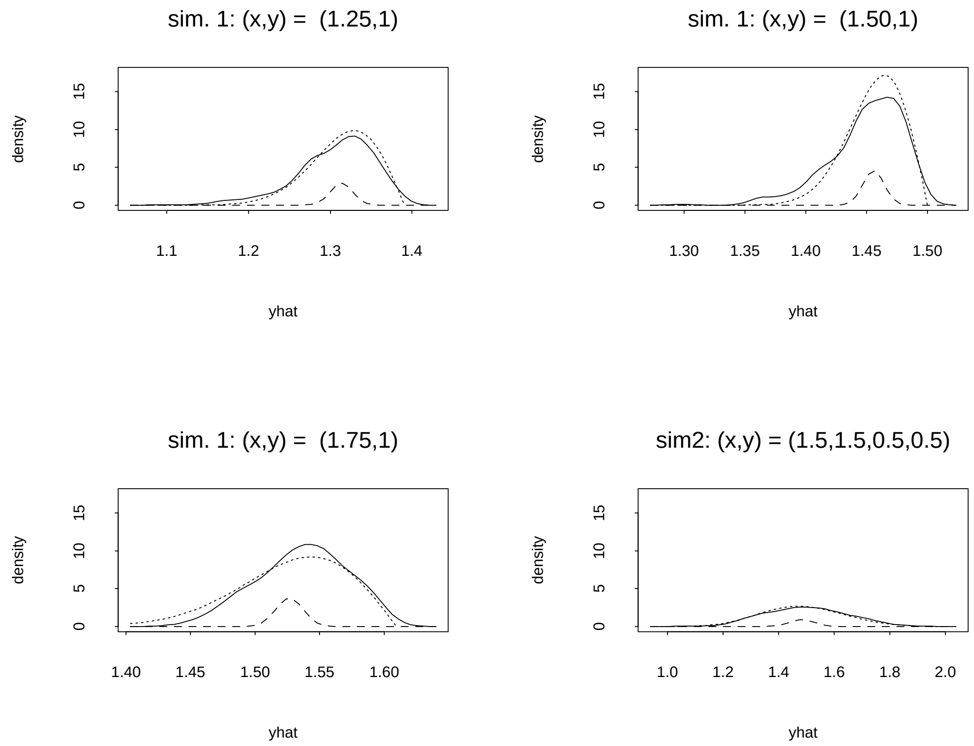

Figure 2: *kernel estimate of the empirical distribution (solid line) compared with the density of a Weibull distribution (doted line) for a sample with 100 observations. The empirical density is estimated with a kernel estimator using an Epanechnikov kernel and a bandwidth according to Silverman's rule of thumb. The kernel divided by* 10 *with the actual bandwidth (dashed line) is plotted to illustrated the degree of smoothing*

here the small sample bias is negligible. A good choice for C must be between 0.1 and 3. $\widehat{\widehat{\mu}}_{NW}$ performs better than $\widehat{\mu}_{NW}$ and for all simulations $C = 1$ is a reasonable value.

Consider that ζ is a scale parameter for the height of $NW(x_0, y_0^{\partial} - \zeta y_0)$. Hence C does not' t depend on the range of the data, respectively we need not to adapt C to the range of observations. Therefore we generally propose $C = 1$ as a rule of thumb. Moreover figure 3 provides a tool to check this choice: We can compare $\widehat{\mu}_{NW}$ and $\widehat{\widehat{\mu}}_{NW}$ for some representative points in the range of observations for $C \in (0, 3]$. If these plots don't look like the plots in the simulations, we have to choose another C.

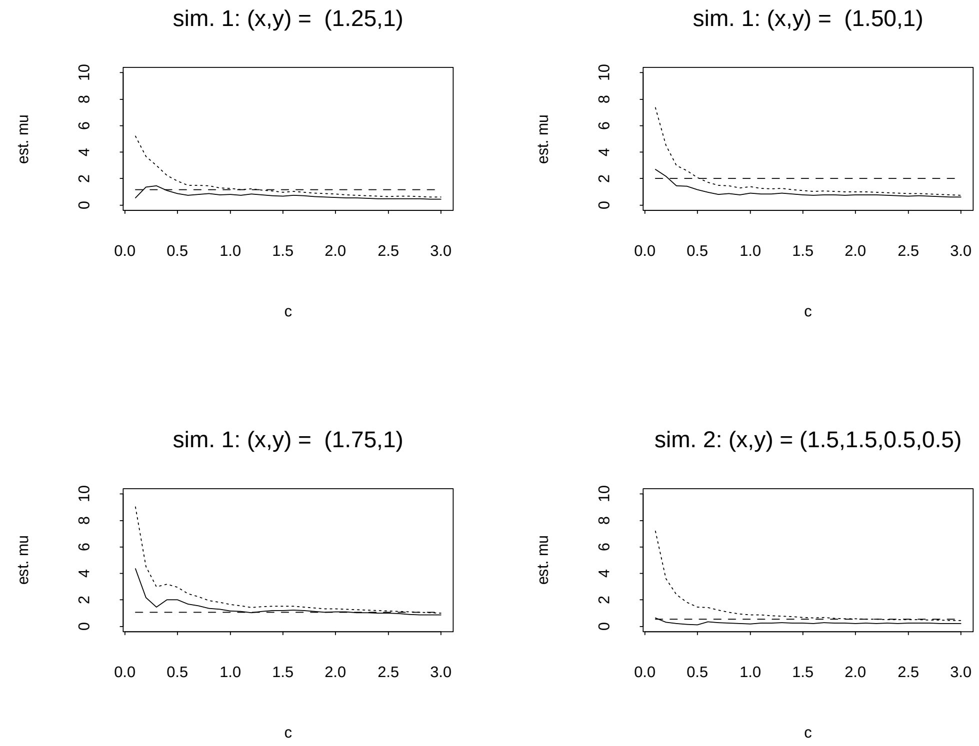

Figure 3: *estimates for μ_{NW} (dashed line) with $\widehat{\mu_{NW}}$ (dotted line) and $\widehat{\widehat{\mu_{NW}}}$ (solid line) for different $\zeta = Cn^{-2/(3(p+q))}$, $C \in (0,3]$; C is plotted on the x axis, μ_{NW}, $\widehat{\mu_{NW}}$ and $\widehat{\widehat{\mu_{NW}}}$ is plotted on the y axis. $\widehat{\widehat{\mu_{NW}}}$ performs better than $\widehat{\mu_{NW}}$ and $C = 1$ yields reasonable estimates for μ_{NW}. In simulation I $\widehat{\widehat{\mu_{NW}}}$ underestimates μ_{NW} for $x_0 = 1.25$ and $x_0 = 1.5$ and it overestimates μ_{NW} at $x_0 = 1.75$; in simulation II μ_{NW} is again underestimated.*

In simulation I $\widehat{\widehat{\mu_{NW}}}$ underestimates μ_{NW} for $x_0 = 1.25$ and $x_0 = 1.5$ and it overestimates μ_{NW} at $x_0 = 1.75$; in simulation II μ_{NW} is again underestimated. This deviation is the opposite of the deviation of a Weibull distribution from the empirical distribution, which we have seen above. Yet, the deviation has the same reason: estimating p_{NW}, we count the observations in $NW(x_0, \widehat{y_0^\partial} - \zeta y_0)$. This region is larger (smaller) than the simplex, if the curve is concave (convex) . Hence our method to estimate μ_{NW} corrects a part of the error of the asymptotic distribution.

(x_0, y_0)	n	ϑ	mean		sd		MSE	
			$\widehat{\vartheta}_0$	$\widehat{\widehat{\vartheta}}_0$	$\widehat{\vartheta}_0$	$\widehat{\widehat{\vartheta}}_0$	$\widehat{\vartheta}_0$	$\widehat{\widehat{\vartheta}}_0$
simulation I								
(1.25,1)	25	1.391	1.205	1.645	0.110	0.572	0.035	0.064
(1.25,1)	50	1.391	1.261	1.475	0.075	0.264	0.017	0.007
(1.25,1)	100	1.391	1.306	1.430	0.047	0.109	0.007	0.002
(1.25,1)	250	1.391	1.338	1.405	0.029	0.042	0.003	0.000
(1.25,1)	500	1.391	1.355	1.400	0.019	0.025	0.001	0.000
(1.5,1)	25	1.500	1.381	1.601	0.079	0.176	0.014	0.010
(1.5,1)	50	1.500	1.428	1.557	0.044	0.066	0.005	0.003
(1.5,1)	100	1.500	1.450	1.528	0.030	0.038	0.003	0.001
(1.5,1)	250	1.500	1.470	1.513	0.016	0.020	0.001	0.000
(1.5,1)	500	1.500	1.479	1.506	0.011	0.013	0.000	0.000
(1.75,1)	25	1.609	1.483	1.657	0.059	0.164	0.016	0.002
(1.75,1)	50	1.609	1.511	1.610	0.042	0.056	0.010	0.000
(1.75,1)	100	1.609	1.536	1.600	0.035	0.043	0.005	0.000
(1.75,1)	250	1.609	1.562	1.600	0.023	0.028	0.002	0.000
(1.75,1)	500	1.609	1.574	1.602	0.018	0.022	0.001	0.000
simulation II								
(1.5,1.5,0.5,0.5)	25	2.003	1.251	2.572	0.215	1.201	0.565	0.325
(1.5,1.5,0.5,0.5)	50	2.003	1.391	2.339	0.189	0.898	0.374	0.113
(1.5,1.5,0.5,0.5)	100	2.003	1.498	2.195	0.143	0.627	0.255	0.037
(1.5,1.5,0.5,0.5)	250	2.003	1.591	2.088	0.111	0.435	0.170	0.007
(1.5,1.5,0.5,0.5)	500	2.003	1.659	2.047	0.092	0.243	0.118	0.002

Table 3: *mean, standard deviation and mean squared error (MSE) for FDH estimates $\widehat{\vartheta}_0$ and bias corrected estimates $\widehat{\widehat{\vartheta}}_0$. Except for simulation I, $x_0 = 1.25$, $n = 25$, $\widehat{\widehat{\vartheta}}_0$ has a smaller bias and MSE. The standard deviation of $\widehat{\widehat{\vartheta}}_0$ is always larger.*

4.1.3 Performance of the bias corrected estimator

In the last step we compare mean, standard deviation and mean squared error (MSE) of the FDH estimator $\widehat{\vartheta}_0$ and the bias corrected estimator $\widehat{\widehat{\vartheta}}_0$. We see numerical expressions is table 3; the values are derived from 500 iterations, μ_{NW} has been estimated using $\widehat{\mu_{NW}}$ and $\zeta = n^{-2/(3(p+q))}$. In all cases, except for simulation I, $x_0 = 1.25$, the bias corrected estimator has a smaller bias and a smaller MSE. The standard deviation is for all estimates larger; this is reasonable, since we have estimated μ_{NW} from the data.

4.2 Real Data Example

We use the NBER Manufacturing Productivity Database (NBER: National Bureau of Econometric Research - Cambridge, MA; USA). It contains annual data from 1958 to 1991 on 450 manufacturing industries in the United States. The data are available in the internet (`http:\\www.nber.org` or `ftp ftp.nber.org`), there is also a description of the data set.

We have chosen the data from 1986 and the following variables:

x^1	total employee compensation
x^2	cost of material input
x^3	energy expenditure
x^4	real capital stock
y	value added

The range of observations is

	min	$z_{0.25}$	$z_{0.5}$	$z_{0.75}$	max	mean
x^1	3.1	193.75	425.65	925.475	16679.9	894.59
x^2	1.8	458.075	1153.95	2394.15	95995.2	2707.13
x^3	0.1	17.9	39.3	99.575	4009.5	118.75
x^4	22	488.525	1032.35	2292.275	62539.8	2582.16
y	5.7	475.275	1078.65	2297.425	34296.4	2301.05

To check the choice of c we look to $\widehat{\mu_{NW}}$ and $\widehat{\widehat{\mu_{NW}}}$ at all points

$$(x, y) = (z_*(x^1), z_*(x^2), z_*(x^3), z_*(x^4), mean(y))$$

where $z_*(x^k)$ denotes the quartiles of x^k (81 points). All plots indicate to choose $c = 1$. (Because there are 81 plots, we do not include them in the paper).

In figure 4 we show estimates and confidence bounds for $\vartheta(X_i, Y_i)$, the FDH estimate (dashed line) which is also the lower confidence bound, the bias corrected estimate (solid line) and the upper confidence bound (dotted line). These triples are sorted according to the bias corrected estimate. On the x axis we see the ranks of the observations, on the y axis we see estimates and confidence bounds for $\vartheta(X_i, Y_i)$.

Interpreting the confidence bounds we must be careful, since we have only pointwise confidence bounds. There is certainly correlation between the estimates.

There is a region from 65 to 195, where the lines are constant. Note that these are the individuals, where the FDH estimate is 1; note furthermore that these estimates have large error terms. There are only a few observations with FDH estimate 1 and small error terms at the left side.

5 Conclusions

If we assume in productivity analysis, that observations are independently drawn from a population, where the support of the distribution is the production set Ψ, than the FDH, FDH efficiency scores and the FDH frontier are objects with random errors.

Under some regularity conditions the estimator for efficiency scores ϑ_0 is asymptotically $W(\mu_{NW} n^{1/(p+q)}, p+q)$, distributed. This implies that good estimates for ϑ_0 require a sample size which exponentially depends on the dimension of in- and output $p+q$. This is the *curse of dimensionality.* Moreover, as $p+q$ increases, the bias dominates the variance of the estimator.

We propose an estimator for the parameter μ_{NW}. This enables us to correct the bias and to estimate confidence bounds. Dealing with real data we can improve the FDH estimator and highlight its random variation. Yet, we only have pointwise confidence bounds and we can certainly assume that the estimates for the efficiency scores are correlated.

Our estimator $\widehat{\mu_{NW}}$ has nice properties, but it could be probably improved. More research is necessary for a completely data driven and more exact choice of ζ. Furthermore in our calculations and simulations, we only consider production plans (x_0, y_0) in the interior of the range of observations. However, in nonparametric statistics we know that something goes wrong for marginal observations. This question has to be investigated, too.

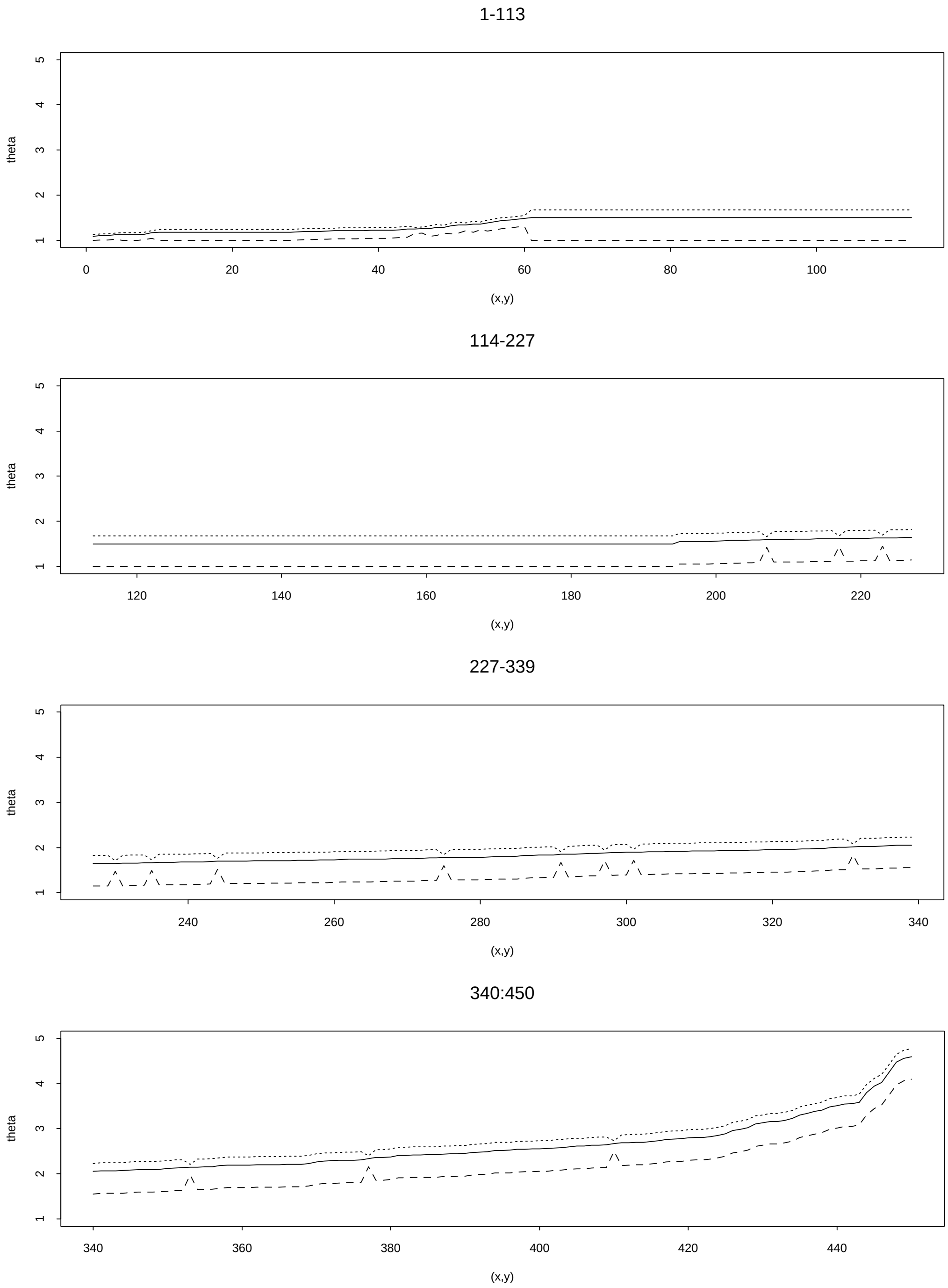

Figure 4: *real data example: lower confidence bound (= FDH estimate) (dashed line), bias corrected estimate (solid line) and upper confidence bound (dotted line) for* $\vartheta(X_i, Y_i)$*, sorted with respect to the bias corrected estimator*

A Lemmas and proofs

Lemma A.1 *(With the notations of chapter 2)*

$$\Psi^k(g_\Psi^k) \ = \ \Psi$$

Proof Consider a pair $(x, y) \in \Psi$. As $y^k \in \{\eta | (x, y^{(k)}(\eta)) \in \Psi\}$, we have $y^k \leq \max\{\eta | (x, y^{(k)}(\eta)) \in \Psi\}$; thus $(x, y) \in \Psi^k(g_\Psi^k)$.

Take on the other hand a pair $(x, y) \in \Psi^k(g_\Psi^k)$. Here we have $y^k \leq g_\Psi^k(x, y^{(k)}) = \max\{\eta | (x, y^{(k)}(\eta) \in \Psi\}$ and hence $y \leq y^{(k)}(g_\Psi^k(x, y^{(k)})) =: y_*$. As Ψ is free disposal and $(x, y_*) \in \Psi$, this implies $(x, y) \in \Psi$.

q.e.d. ■

Definition A.2 (North West Parameter)

$$\begin{aligned}
f_0 &:= f(x_0, y_0^\partial) \\
g_1 &:= \prod_{k=1}^{p} \frac{\partial}{\partial x^k} g(x, y^{(q)})|_{(x,y^{(q)})=(x_0,y_0^{\partial(q)})} \prod_{l=1}^{q-1} \frac{\partial}{\partial y^l} g(x, y^{(q)})|_{(x,y^{(q)})=(x_0,y_0^{\partial(q)})} \\
\overline{y}_0 &:= \sum_{l=1}^{q-1} |\frac{\partial}{\partial y^l} g(x, y^{(q)})|_{(x,y^{(q)})=(x_0,y_0^{\partial(q)})} |\, y_0^k \ + \ y_0^q \\
\mu_{NW} &:= \Big(\frac{f_0}{|g_1|} \frac{\overline{y}_0^{p+q}}{(p+q)!} \Big)^{\frac{1}{p+q}}
\end{aligned}$$

Lemma A.3 (North West Lemma) *Assume A II, A III, and consider* $\zeta \to 0$. *Then it holds:*

$$p_{NW}(\zeta) \ = \ (\mu_{NW}\zeta)^{p+q} + O(\zeta^{p+q+1}).$$

Proof A II implies $f(x_0, y_0^\partial - \zeta y_0) = f_0 + O(\zeta)$ and

$$p_{NW}(\zeta) \ = \ (f_0 + O(\zeta))\lambda[\,\{NW(x_0, y_0^\partial - \zeta y_0)\} \cap \Psi\,] \tag{10}$$

where λ denotes the Lebesgue measure. As we know that a pair (x, y) is in Ψ if and only if $y^q \leq g(x, y^{(q)})$, we can write

$$\lambda[NW(x_0, y_0^\partial - \zeta y_0) \cap \Psi] = \int_{(-\infty,x_0]\times[y_0^\partial - \zeta y_0,\infty)} I_{\{y^q \leq g(x,y^{(q)})\}} dx dy \tag{11}$$

Using a first order approximation for g in a region around $(x_0, y_0^{\partial(q)})$, we derive an approximation for (11): let $d_0(x, y) := ((x - x_0), (y - y_0^\partial)^{(q)})$ and

$\nabla g_0 := \nabla g(x_0, y_0^{\partial(q)})$. Consider $g(x_0, y_0^{\partial(q)}) = y_0^{\partial q}$, and $\|d_0(x,y)\| = O(\zeta)$ for $(x,y) \in NW(x_0, y_0^\partial - \zeta y_0) \cap \Psi$. A III yields $g(x, y^{(q)}) = y_0^{\partial q} + \nabla g_0^t d_0(x,y) + O(\zeta^2)$. Therefore

$$\lambda[NW(x_0, y_0^\partial - \zeta y_0) \cap \Psi] = \int_{(-\infty, x_0] \times [y_0^\partial - \zeta y_0, \infty)} I_{\{\nabla g_0^t d_0(x,y) \geq y^q - y_0^{\partial q} + O(\zeta^2)\}} dx dy$$

In order to simplify this term we move $(x_0, y_0^\partial - \zeta y_0)$ to the origin, change the orientation of all x variables, and rescale with the absolute value of all partial derivatives. In particular we transform all x^k to $\tilde{x}^k := \frac{\partial}{\partial x^k} g(x_0, y_0^{\partial(q)})(x_0^k - x^k)$ for $1 \leq k \leq p$, y^l to $\tilde{y}^l = |\frac{\partial}{\partial y^l} g(x_0, y_0^{\partial(q)})|(y^l - y_0^{\partial l} + \zeta y_0^l)$ for $1 \leq l \leq q-1$ and y^q to $\tilde{y}^q = y^q - y_0^{\partial q} + \zeta y_0^q$. We obtain

$$\lambda[NW(x_0, y_0^\partial - \zeta y_0) \cap \Psi] = |g_1|^{-1} \int_{[0,\infty)^{p+q}} I_{\{\sum_{k=1}^p \tilde{x}^k + \sum_{l=1}^q \tilde{y}^l \leq \zeta(\overline{y}_0 + O(\zeta))\}} d\tilde{x} d\tilde{y}$$

The integral is the volume of a $p+q$ dimensional simplex with edges of the length $\zeta(\overline{y}_0 + O(\zeta))$. Together with equation (10) this shows the lemma.

q.e.d. ■

Proof of Theorem 3.3 Let $X_n := n^{1/(p+q)}(\vartheta_0 - \widehat{\vartheta_0})$. We show X_n^r is uniformly integrable, i. e. $\lim_{c\to\infty} \sup_n E[X_n^r I_{\{X_n \geq c\}}] = 0$; Since the distribution of X_n weakly converges to a $W(\mu_{NW}, p+q)$ distribution (theorem 3.1) this implies the theorem. (cf. Serfling [11], chapter 1.4, theorem A, p 14.)

Consider

$$\tilde{p}(\zeta) := \begin{cases} p_{NW}(\zeta)/\zeta^{p+q} & \zeta \in (0, \vartheta_0] \\ \mu_{NW}^{p+q} & \zeta = 0 \end{cases}$$

From lemma A.3 we derive that $\tilde{p}$ is a positive continuous function on $[0, \vartheta_0]$ which has a positive minimum $\alpha := \min_{\zeta \in [0,\vartheta_0]} \tilde{p}(\zeta) > 0$. Hence it holds $p_{NW}(\zeta) \geq \alpha \zeta^{p+q}$. Furthermore consider $(1-x)^n \leq e^{-nx}$ for $0 \leq x \leq 1$. Using $E[X^r I_{\{X \geq c\}}] = c^r(1 - P[X \leq c]) + r \int_c^\infty z^{r-1}(1 - P[X \leq z]) dz$, we get

$$\begin{aligned} & E[X_n^r I_{\{X_n \geq c\}}] \\ &= c^r(1 - p_{NW}(n^{-1/(p+q)}c))^n + r \int_c^{\vartheta n^{1/(p+q)}} z^{r-1}(1 - p_{NW}(n^{-1/(p+q)}z))^n dz \\ &\leq c^r \exp[-\alpha c^{p+q}] + r \int_c^\infty z^{r-1} \exp[-\alpha z^{p+q}] dz \end{aligned}$$

Since $\int_0^\infty z^{r-1} \exp[-\alpha z^{p+q}] dz < \infty$, this implies uniform integrability of X_n.

q.e.d. ■

Lemma A.4 *Assume A I, A II and A III and consider* $\zeta \to 0$, *such that* $n^{-1/(p+q)}\zeta^{-1} \to 0$. *Mean and variance of* $\widehat{p_{NW}}(\zeta)$ *are:*

$$\begin{aligned} E[\widehat{p_{NW}}(\zeta)] &= \left(\mu_{NW}\zeta + c_1(n-1)^{-1/(p+q)}\right)^{p+q} \\ &\qquad + o(\zeta^{p+q-1}n^{-1/(p+q)}) + O(\zeta^{p+q+1}) \\ var[\widehat{p_{NW}}(\zeta)] &= O(\zeta^{2(p+q)-2}n^{-2/(p+q)}) + O(\zeta^{2(p+q)+1}) \end{aligned}$$

where $c_1 = \Gamma((p+q+1)/(p+q))$

Proof Let $\widehat{\vartheta_0^{-i}}$ denote the "leave one out estimator", i. e. the efficiency which is estimated from the $(n-1)$ observations leaving out the ith pair of in- and output. We may write $\widehat{p_{NW}}(\zeta)$ as follows

$$\widehat{p_{NW}}(\zeta) = \frac{1}{n}\sum_{i=1}^{n} I_{[0,x_0]\times[\widehat{\vartheta_0^{-i}}y_0-\zeta y_0,\infty)}(X_i,Y_i) - I_{[0,x_0]\times[\widehat{\vartheta_0^{-i}}y_0-\zeta y_0,\widehat{\vartheta_0}y_0-\zeta y_0)}(X_i,Y_i)$$

The second term in the summation, $I_{[0,x_0]\times[\widehat{\vartheta_0^{-i}}y_0-\zeta y_0,\widehat{\vartheta_0}y_0-\zeta y_0)}(X_i,Y_i)$, corrects the difference between $\widehat{\vartheta_0}$ and $\widehat{\vartheta_0^{-i}}$. Note that this term is zero, since $\widehat{\vartheta_0^{-i}} \neq \widehat{\vartheta_0}$ if and only if "Y_i is the most efficient competitor", i. e. $\min_{1\le k\le q} Y_i^k/y_0^k = \widehat{\vartheta_0}$; in this case we have $Y_i > \widehat{\vartheta_0}y_0 - \zeta y_0$, especially $Y_i \notin [(\widehat{\vartheta_0^{-i}}y_0 - \zeta y_0, \widehat{\vartheta_0}y_0 - \zeta y_0)$.

The mean of the first indicator function, $I_{[0,x_0]\times[\widehat{\vartheta_0^{-i}}y_0-\zeta y_0,\infty)}(X_i,Y_i)$, is

$$\begin{aligned} &E\Big[\,P\Big[(X_1,Y_1) \in NW(x_0,(\widehat{\vartheta_0^{-1}}-\zeta)y_0)\,\Big|\,\{(X_i,Y_i); i\ge 2\}\Big]\,\Big] \\ &= E\Big[\mu_{NW}^{p+q}(\zeta+\vartheta_0-\widehat{\vartheta_0^{-1}})^{p+q} + O((\zeta+\vartheta_0-\widehat{\vartheta_0^{-1}})^{p+q+1})\Big] \\ &= (\mu_{NW}\zeta)^{p+q} + (p+q)(\mu_{NW}\zeta)^{p+q-1}m_1(n-1)^{-\frac{1}{p+q}} \\ &\qquad + O(\zeta^{p+q-2}n^{-2/(p+q)}) + O(\zeta^{p+q+1}) \end{aligned}$$

where m_1 denotes the first moment of $\mu_{NW}n^{1/(p+q)}(\vartheta_0 - \widehat{\vartheta_0})$. Because $m_1 = \Gamma((p+q+1)/(p+q)) + o(1)$ (theorem 3.3), this shows the mean of $\widehat{p_{NW}}(\zeta)$.

For the variance consider

$$\begin{aligned} &\mathrm{var}[\widehat{p_{NW}}(\zeta)] \\ &= \frac{1}{n}\mathrm{var}[I_{\{NW(x_0,(\widehat{\vartheta_0^{-1}}-\zeta)y_0\}}(X_1,Y_1)] \\ &\quad + \frac{1}{n^2}\sum_{i\neq j}\mathrm{cov}[I_{NW(x_0,(\widehat{\vartheta_0^{-i}}-\zeta)y_0)}(X_i,Y_i), I_{NW(x_0,(\widehat{\vartheta_0^{-j}}-\zeta)y_0)}(X_j,Y_j)]. \end{aligned}$$

Let $\widehat{\vartheta_0^{-i,j}}$ denote the "leave two out" estimator, leaving out the ith and jth observations. Then by the same tick as above, we get

$$\begin{aligned} & I_{NW(x_0,(\widehat{\vartheta_0^{-i}}-\zeta)y_0)}(X_i,Y_i)\, I_{NW(x_0,(\widehat{\vartheta_0^{-j}}-\zeta)y_0)}(X_j,Y_j) \\ & \qquad = I_{NW(x_0,(\widehat{\vartheta_0^{-i,j}}-\zeta)y_0)}(X_i,Y_i)\, I_{NW(x_0,(\widehat{\vartheta_0^{-i,j}}-\zeta)y_0)}(X_j,Y_j) \end{aligned}$$

Now, using conditioning arguments,

$$\begin{aligned} & E[I_{NW(x_0,(\widehat{\vartheta_0^{-i,j}}-\zeta)y_0)}(X_i,Y_i) I_{NW(x_0,(\widehat{\vartheta_0^{-i,j}}-\zeta)y_0)}(X_j,Y_j)] \\ & \quad = E\Big[P\Big[(X_1,Y_1) \in \text{ NW of } (x_0,(\widehat{\vartheta_0^{-1,2}}-\zeta)y_0) \,\Big|\, \{(X_i,Y_i); i\geq 3\}\Big] \\ & \qquad\qquad \cdot P\Big[(X_2,Y_2) \in \text{ NW of } (x_0,(\widehat{\vartheta_0^{-1,2}}-\zeta)y_0) \,\Big|\, \{(X_i,Y_i); i\geq 3\}\Big]\Big] \\ & \quad = \mu_{NW}^{2(p+q)}\zeta^{2(p+q)} + 2(p+q)(\mu_{NW}\zeta)^{2(p+q)-1} m_1 (n-2)^{-1/(p+q)} \\ & \qquad\qquad + O(\zeta^{2(p+q)-2} n^{-2/(p+q)}) + O(\zeta^{2(p+q)+1}). \end{aligned}$$

Hence the covariances are

$$\begin{aligned} & \operatorname{cov}[I_{NW(x_0,(\widehat{\vartheta_0^{-i}}-\zeta)y_0)}(X_i,Y_i), I_{NW(x_0,\widehat{\vartheta_0^{-j}}-\zeta)y_0}(X_j,Y_j)] \\ & \qquad = \mu_{NW}^{2(p+q)}(\zeta^{2(p+q)} + 2(p+q)\zeta^{2(p+q)-1} m_1 \mu_{NW}^{-1}(n-2)^{-1/(p+q)}) \\ & \qquad\quad - \mu_{NW}^{2(p+q)}(\zeta^{2(p+q)} + 2(p+q)\zeta^{2(p+q)-1} m_1 \mu_{NW}^{-1}(n-1)^{-1/(p+q)}) \\ & \qquad\qquad + O(\zeta^{2(p+q)-2} n^{-2/(p+q)}) + O(\zeta^{2(p+q)+1}) \\ & \qquad = O(\zeta^{2(p+q)-2} n^{-2/(p+q)}) + O(\zeta^{2(p+q)+1}) \end{aligned}$$

Since $\frac{1}{n}\operatorname{var}[I_{\{NW(x_0,(\widehat{\vartheta_0^{-1}}-\zeta)y_0\}}(X_1,Y_1)] = O(\zeta^{p+q}n^{-1})$ goes to zero faster than the covariances, the variance is

$$\operatorname{var}[\widehat{p_{NW}}(\zeta)] \;=\; O(\zeta^{2(p+q)-2} n^{-2/(p+q)}) + O(\zeta^{2(p+q)+1})$$

q.e.d. ■

Proof of Theorem 3.4 Let $\mu(x) := \zeta^{-1}x^{1/(p+q)}$; we have $\widehat{\mu_{NW}} = \mu(\widehat{p_{NW}}(\zeta))$.

bias: We show

$$E[\mu(\widehat{p_{NW}}(\zeta))] \;=\; \mu(E[\widehat{p_{NW}}(\zeta)]) + O(\zeta^{-2}n^{-2/(p+q)}) + O(\zeta) \tag{12}$$

This implies the bias, since we derive from lemma A.4 $\mu(E[\widehat{p_{NW}}(\zeta)]) \;=\; \mu_{NW} + c_1\zeta^{-1}(n-1)^{-1/(p+q)} + o(\zeta^{-1}n^{-1/(p+q)}) + O(\zeta)$.

Consider for $0 < c < 1$

$$
\begin{aligned}
E[\mu(\widehat{p_{NW}}(\zeta))] &- \mu(E[\widehat{p_{NW}}(\zeta)]) \\
&= E[\Big(\mu(\widehat{p_{NW}}(\zeta)) - \mu(E[\widehat{p_{NW}}(\zeta)])\Big) I_{\widehat{p_{NW}}(\zeta)\geq(1-c)E[\widehat{p_{NW}}(\zeta)]}] \\
&\quad + E[(\mu(\widehat{p_{NW}}(\zeta)) - \mu(E[\widehat{p_{NW}}(\zeta)]))I_{\widehat{p_{NW}}(\zeta)<(1-c)E[\widehat{p_{NW}}(\zeta)]}] \qquad (13)
\end{aligned}
$$

Let A and B denote the two summands in the right hand side of (13).

We apply a second order approximation to A:

$$
\begin{aligned}
A &= \mu'(E[\widehat{p_{NW}}(\zeta)])E\Big[(\widehat{p_{NW}}(\zeta) - E[\widehat{p_{NW}}(\zeta)])\, I_{\widehat{p_{NW}}(\zeta)\geq(1-c)E[\widehat{p_{NW}}(\zeta)]}\Big] \\
&\quad + E\Big[\mu''(\tilde{p})(\widehat{p_{NW}}(\zeta) - E[\widehat{p_{NW}}(\zeta)])^2\, I_{\widehat{p_{NW}}(\zeta)\geq(1-c)E[\widehat{p_{NW}}(\zeta)]}\Big]
\end{aligned}
$$

where $\tilde{p} \geq (1-c)E[\widehat{p_{NW}}(\zeta)]$. For the first term we use $\mu'(E[\widehat{p_{NW}}(\zeta)]) = O(\zeta^{-(p+q)})$, $E[\widehat{p_{NW}}(\zeta) - E[\widehat{p_{NW}}(\zeta)]] = 0$, Hölder's inequality and Chebychev's inequality:

$$
\begin{aligned}
\mu'(E[\widehat{p_{NW}}(\zeta)])&E\Big[(\widehat{p_{NW}}(\zeta) - E[\widehat{p_{NW}}(\zeta)])\, I_{\widehat{p_{NW}}(\zeta)\geq(1-c)E[\widehat{p_{NW}}(\zeta)]}\Big] \\
&= O(\zeta^{-(p+q)})E\Big[(E[\widehat{p_{NW}}(\zeta)] - \widehat{p_{NW}}(\zeta))I_{\widehat{p_{NW}}(\zeta)<(1-c)E[\widehat{p_{NW}}(\zeta)]}\Big] \\
&\leq O(\zeta^{-(p+q)}\mathrm{var}[\widehat{p_{NW}}(\zeta)]^{1/2}P\Big[E[\widehat{p_{NW}}(\zeta)] - \widehat{p_{NW}}(\zeta) > cE[\widehat{p_{NW}}(\zeta)]\Big]^{1/2} \\
&\leq O(\zeta^{-(p+q)})\mathrm{var}[\widehat{p_{NW}}(\zeta)]/cE[\widehat{p_{NW}}(\zeta)] \;=\; O(\zeta^{-2}n^{-2/(p+q)})
\end{aligned}
$$

For the second term we use $|\mu''(\tilde{p})| \leq |\mu''((1-c)E[\widehat{p_{NW}}(\zeta)])|$; hence we get

$$
\begin{aligned}
|A| &\leq O(\zeta^{-2}n^{-2/(p+q)}) + \Big|\mu''\Big((1-c)E[\widehat{p_{NW}}(\zeta)]\Big)\Big|\,\mathrm{var}[\widehat{p_{NW}}(\zeta)] \\
&= O(\zeta^{-2}n^{-2/(p+q)}) + O(\zeta)
\end{aligned}
$$

For B we use Chebychev's inequality: Consider that μ is monotonically increasing. Hence $\widehat{p_{NW}}(\zeta) < (1-c)E[\widehat{p_{NW}}(\zeta)] < E[\widehat{p_{NW}}(\zeta)]$ implies $\mu(\widehat{p_{NW}}(\zeta)) \leq \mu(E[\widehat{p_{NW}}(\zeta)])$. We have

$$
|B| \leq 2\mu(E[\widehat{p_{NW}}(\zeta)])\,\mathrm{var}[\widehat{p_{NW}}(\zeta)]/(cE[\widehat{p_{NW}}(\zeta)])^2 \;=\; O(\zeta^{-2}n^{-2/(p+q)}) + O(\zeta)
$$

variance: For the variance we use the same trick: We have

$$
\begin{aligned}
E[(\mu(\widehat{p_{NW}}(\zeta)) &- \mu(E[\widehat{p_{NW}}(\zeta)]))^2] \\
&= E\Big[\Big(\mu(\widehat{p_{NW}}(\zeta)) - \mu(E[\widehat{p_{NW}}(\zeta)])\Big)^2 I_{\widehat{p_{NW}}(\zeta)\geq(1-c)E[\widehat{p_{NW}}(\zeta)]}\Big] \\
&\quad + E\Big[(\mu(\widehat{p_{NW}}(\zeta)) - \mu(E[\widehat{p_{NW}}(\zeta)]))^2 I_{\widehat{p_{NW}}(\zeta)<(1-c)E[\widehat{p_{NW}}(\zeta)]}\Big] \qquad (14)
\end{aligned}
$$

Let A' and B' denote the two summands in the right hand side of (14).

For A' we apply a first order approximation: Consider that $(\mu'(\cdot))^2$ is monotonically decreasing and $E[\widehat{p_{NW}}(\zeta)] = O(\zeta^{p+q})$. We have

$$|A'| \;\leq\; \Big(\mu'\Big((1-c)E[\widehat{p_{NW}}(\zeta)]\Big)\Big)^2 \operatorname{var}[\widehat{p_{NW}}(\zeta)] \;=\; O(\zeta^{-2}n^{-2/(p+q)}) + O(\zeta)$$

For B' we use Chebychev's inequality again:

$$|B'| \leq 4\mu(E[\widehat{p_{NW}}(\zeta)])^2 \operatorname{var}[\widehat{p_{NW}}(\zeta)]/(cE[\widehat{p_{NW}}(\zeta)])^2 = O(\zeta^{-2}n^{-2/(p+q)}) + O(\zeta)$$

q.e.d. ■

References

[1] Afriat, S. (1972) Efficiency Estimation of Production Functions. *International Economic Review* 13, pp. 568 - 598.

[2] Banker, R.D. (1993) Maximum likelihood, consistency and data envelopment analysis: a statistical foundation. *Management Science* 39, 10, pp. 1265 - 1273.

[3] Charnes, A., Cooper, Rhodes, E. (1978) Measuring the inefficiency of decision making units. *European Journal of Operations Research* 2, pp. 429 - 444.

[4] Deprins, D., Simar, L., Tulkens, H. (1984) Measuring Labor Efficiency in Post Offices, in M. Marchand, P. Pestieau, H. Tulkens (eds): *The Performance of Public Enterprises: Concepts and Measurements*, pp. 243 - 267. Amsterdam: North - Holland.

[5] Färe, R., Grosskopf, S., Lovell, C. A. K. (1985) *The Measurment of Efficiency of Production*. Studies in Productivity Analysis. Boston Dordrecht Lancaster: Kluwer - Nijhoff Publishing.

[6] Gijbels, I., Mammen, E., Park, B.U., Simar, L. (1996) On Estimation of monotone and concave frontier functions. *Discussion Paper DP9611, Institute de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.* (ftp: stat1ux.stat.ucl.ac.be; login: anonymous; cd pub/papers/dp)

[7] Farrell, M.J. (1957) The measurement of productive efficiency. *Journal of the Royal Statistical Society, Series A* 120, pp. 253 - 281

[8] Kneip, A., Park, B.U., Simar, L. (1996) A note on the convergence of nonparametric DEA efficiency measures. *submitted to Econometric Theory*

[9] Korostelev, A., Simar, L., Tsybakov, A.B. (1995a) Efficient estimation of monotone boundaries. *The Annals of Statistics* 23, pp. 476-489

[10] Korostelev, A., Simar, L., Tsybakov, A.B. (1995b) On the estimation of monotone and convex boundaries. *Pub. Inst. Stat. Univ Paries* XXXIX, 1, pp. 3-18

[11] Serfling, R. J. (1980) *Approximation Theorems of Mathematical Statistics*. Wiley Series in Probability and Mathematical Statistics. New York, Chichester, Brisbane, Toronto: John Wiley & Sons

[12] Simar, L., Wilson, P. (1996) Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. manuscript to appear in *Management Science*