Estimating Efficiency Offset between Two Groups of Decision-Making Units

Karel Macek[1]

Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic,
Pod Vodárenskou věží 4, Prague 8, 180 00, Czech Republic, karel.macek@utia.cas.cz

Keywords: Data Envelopment Analysis; Local Regression; Efficiency Comparison; Interval Estimation

Abstract. The comparison of two groups of decision-making units (DMUs) has been already subject of scientific reflection. So far, some statistical tests have been developed. This article addresses estimating the difference between expected outputs of two groups of DMUs. In contrast to other efficiency evaluation methods, this publication focuses on quantitative assessment of this difference, not on the hypothesis testing. The article focuses on single output DMUs and the designed statistical tests are examined on various simulated data sets as well as on one real-world example. Some of them stem from the data envelopment analysis, others are related to the local regression.

1 Introduction

Efficiency evaluation of decision making units (DMUs) has attracted the attention since 1957 [1]. Later on, a comparison of two or more groups of decision making units extended the basic framework of individual efficiencies.

One of the most important efforts was [2] where the statistical foundation of the data envelopment analysis (DEA) is introduced as well as a statistical test dedicated to comparison of efficiency offgroups in two groups of DMUs. This work was extended to 5 alternatives in [3].

This work offers a way how to examine the expected offset between two groups of DMUs. Some of them are based on simple tests from [3] while others apply local regression as a benchmark [4, 5]. The two groups of DMUs could correspond to farms in with green vs. classic approach or branches of two banks. Finally, those two groups could correspond to operation of two systems (e.g. factories) or one system with two different configurations (e.g. one factory using two different production programs). These examples are summarized in Table 1. In this table, the three first rows focus on the qualitative assessment which was mainly based on the hypothesis testing. This can be used for determination if there are some differences in the efficiency between the two groups of the DMUs. The last row in the table is our contribution, where it is also important to quantify the efficiency in an explicit way. If a new technology, a decision support system, or an outsourced service are paid and bought, it is important to quantify the value added by a new approach. In case of so called performance contracts [1], this quantification

[1] See e.g. http://www1.eere.energy.gov/femp/financing/espcs.html

The Proceedings of ECML/PKDD 2013 Workshop
Scalable Decision Making: Uncertainty, Imperfection, Deliberation (SCALE)
September 23, 2013, Prague, Czech Republic

determines the payments to the provider of the new approach and improves the competition between providers.

Table 1. Examples of offset measurement of two groups of DMUs

DMUs	Inputs	Output	Groups differ by	Purpose of offset evaluation
Family farmers [6]	cultivated area, working days	net income	traditional vs. green farming	is green farming sufficient for the families
Coffee retailers [7]	costs of goods sold; sales, general, and administrative expenses; depreciation/amortization	revenue	fair-trade vs. others	competitiveness impact of socially responsible sourcing
Universities [8]	staff; non-personnel expenditures	students, publications, third party funds	German vs. Swiss	evaluation of EU initiated reforms
Operation days of a building [5]	daily average of the ambient temperature	power consumption	original vs. new controller of the HVAC system	evaluation of savings achieved by the new technology

At more general level, the efficiency evaluation of DMUs is of high importance in large-scale distributed systems where the quality of different decision making approaches has to be evaluated. This can lead to propagation of positive experience within a DMU network.
The text is organized as follows: Sect. 2 introduces used notation. The notation is used in Sect. 3 for the problem formulation, i.e. the estimation of output offset between two groups of DMUs. Consequently, some estimates of the offset are provided in Sect. 4. Those estimates are examined on both simulated and real data in Sect. 5. The text is concluded in Sect. 6 by a short summary.

2 Notation

Before we will define specific notation for the addressed domain, we introduce some general notation. We will use $\mathbb{N}$ for natural numbers, $\mathbb{R}$ for the set of real numbers and $\mathbb{R}^N$ for N dimensional real vectors. The equality by definition is denoted by $\equiv$. The conditioned probability density function are denoted as $f(\cdot|\cdot)$ and are distinguished by their arguments. The conditioned expected value is defined as $\mathcal{E}[a|b] \equiv \int a f(a|b) \mathrm{d}a$.

Let us consider two groups of DMUs. Each DMU transforms the inputs to output and each group can use different mechanisms for this transformation. Mathematically, each DMU has a single output $y \in \mathbb{R}$ and a vector input $\mathbf{x} \in \mathbb{R}^m$, having dimension $m \in \mathbb{N}$. For the comparison, we have data in form $D = (\mathbf{x}^{(i)}, y_i, k_i)_{i=1}^n$ where i are indices of data , $\mathbf{x}^{(i)}$ denotes ith vector with components $\mathbf{x}_j^{(i)}$, and $k_i \in \{1,2\}$ is an index of the group. We assume that the DMUs within each group are homogeneous, i.e. the input-output transformation is described by the probability density function $f(y|\mathbf{x},k)$. This dependency can be modeled using a reference r output and noise terms, i.e.:

$$y = r(\mathbf{x}) + u_k \tag{1}$$

where $r : \mathbb{R}^m \to \mathbb{R}$ and u_k is a general noise term, not necessarily zero-mean. This model assumes that the noise u_k is dependent on the group, but not on the inputs. The tests introduced in [3] do this assumption which is from our point of view not very realistic. In the real situations the output variance might depend on the inputs. Typically, the more input, the higher variance. Therefore, we introduce also an alternative model instead of (1)

$$y = r(\mathbf{x}) + v_k(\mathbf{x}). \tag{2}$$

where v_k is a noise, depending on $\mathbf{x}$.

Furthermore, we assume that each DMU operates under different conditions and using different inputs. Thus each group has its typical inputs and conditions. Therefore we assume the inputs to have pdf $f(\mathbf{x}|k)$. Finally, we introduce the probability that a randomly selected DMU will belong to the first group $\rho \equiv \mathbf{P}(k=1)$. Then the marginal pdf of $\mathbf{x}$ is

$$f(\mathbf{x}) = \rho f(\mathbf{x}|k=1) + (1-\rho) f(\mathbf{x}|k=2). \tag{3}$$

3 Problem Formulation

Consider we have a given input $\tilde{\mathbf{x}} \in \mathbb{R}^m$. We let one DMU from both groups transform this input. First, the expected difference of outputs for given $\tilde{\mathbf{x}}$ is:

$$\delta(\mathbf{x}) \equiv \mathcal{E}[y|\tilde{\mathbf{x}}, k=2] - \mathcal{E}[y|\tilde{\mathbf{x}}, k=1]. \tag{4}$$

Next, the expected average difference equals:

$$\Delta \equiv \mathcal{E}[\delta(\mathbf{x})] = \int_{\mathbb{R}^m} \delta(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}. \tag{5}$$

Note that in case of independence of noise on the input (1), it holds

$$\Delta = \delta(\mathbf{x}) \qquad \forall \mathbf{x} \in \mathbb{R}^m. \tag{6}$$

This proposition can be proved by application of the definition in (1), definitions of δ and Δ. The application of the additivity to (4) leads to elimination of the x dependent parts. A formal proof is beyond the scope of this text. We use the simpler model (1 because we need to know the analytical value of Δ which are carried out in Sect. 5. In case of (2), we it seems to be necessary to approach the real value of Δ using intensive Monte-Carlo simulations.

The problem addressed in this text consists in estimating $\delta(\mathbf{x})$ and Δ from available data. This estimation of the offset between two groups of DMUs can be of three forms: (i) a point estimate $\hat{\Delta}$, (ii) an interval estimate $(\hat{\Delta}_{\min}, \hat{\Delta}_{\max})$, or (iii) a posterior pdf $f(\Delta|D)$ where D are the available data.

4 Considered Estimates

In this Sect. we provide information on the considered estimates of Δ. First, we will introduce the benchmarking models that are data-driven and have minimal assumptions about the structure of (1) and (2). Note we work with a common benchmark for data from both groups[2]. Consequently, we will describe the algorithm for the estimation of Δ where the benchmarking models are used at the first step.

DEA Benchmarking. As proposed in [3], the reference r which has been introduced in (1-2) can be estimated as follows, corresponding to the BCC[3] model [9]:

$$\hat{r}_{\mathrm{bcc}}(\tilde{\mathbf{x}}) = \max\{\phi| \tag{7}$$

$$\sum_{i=1}^{n} \lambda_i y_i = \phi; \tag{8}$$

$$\sum_{i=1}^{n} \lambda_i \mathbf{x}_j^{(i)} \leq \tilde{\mathbf{x}}_j, \forall j = 1, \dots m; \tag{9}$$

$$\sum_{i=1}^{n} \lambda_i = 1; \tag{10}$$

$$\lambda_i \geq 0, \quad \forall i = 1 \dots n\} \tag{11}$$

Let us interpret this reference briefly. For a given $\tilde{\mathbf{x}}$, we are looking for a maximal combined output (8) while the combined inputs are limited by the given one (9). The allowed combinations are convex as stated in conditions on $\lambda_1 \dots \lambda_n$ in (10), (11).

This estimate can be calculated by solving a linear programming problem. We mention also usual modification of the basic DEA apprach. First, we consider the

[2] It is possible also the creation two benchmarks, but this is not addressed in this article. Elsewhere [5], we clarify the motivation for a common benchmark for both groups carefully.

[3] BCC stands for Barker, Charnes, and Cooper who were authors of this model.

FDH[4] model [10]

$$\hat{r}_{\mathrm{fdh}}(\tilde{\mathbf{x}}) = \max\{\phi| \tag{12}$$

$$\sum_{i=1}^{n} \lambda_i y_i = \phi; \tag{13}$$

$$\sum_{i=1}^{n} \lambda_i \mathbf{x}_j^{(i)} \leq \tilde{\mathbf{x}}_j, \forall j = 1, \dots m; \tag{14}$$

$$\sum_{i=1}^{n} \lambda_i = 1; \tag{15}$$

$$\lambda_i \in \{0,1\}\} \tag{16}$$

Then also the basic DEA model denoted as CCR[5] [11]

$$\hat{r}_{\mathrm{ccr}}(\tilde{\mathbf{x}}) = \max\{\phi| \tag{17}$$

$$\sum_{i=1}^{n} \lambda_i y_i = \phi; \tag{18}$$

$$\sum_{i=1}^{n} \lambda_i \mathbf{x}_j^{(i)} \leq \tilde{\mathbf{x}}_j, \forall j = 1, \dots m; \tag{19}$$

$$\lambda_i \geq 0\} \tag{20}$$

The value of the benchmark can be calculated for each $x \in \mathbb{R}^{N_x}$ using a linear programming procedure where ϕ stands for the objective function and the equalities and inequalities for constrains. The DEA approaches differ each other with respect to the conditions on λ_i, as it can be seen in (10), (11), (15), (16), and (20). One can see that conditions for FDH are the most strict while for the CCR are the less strict. Thus, the FDH will rate more units efficient than BCC or CCR.

Local Regression Benchmarking. Another way how to construct a benchmark is a local polynomial regression [4] - abbreviated as LPR - of the degree p. We will use only the single input version, i.e. $m = 1$:

$$\hat{r}_{\mathrm{lpr}}(\tilde{\mathbf{x}}) = \sum_{i=1}^{n} \lambda_i(\tilde{\mathbf{x}}) y_i \tag{21}$$

where the vector $\lambda(\mathbf{x})^T = (\lambda_1, \lambda_2, \dots, \lambda_n)$ is calculated as

$$\lambda(\tilde{\mathbf{x}})^T = \mathbf{e}^T (X_{\tilde{\mathbf{x}}}^T W_{\tilde{\mathbf{x}}} X_{\tilde{\mathbf{x}}})^{-1} X_{\tilde{\mathbf{x}}}^T W_{\tilde{\mathbf{x}}} \tag{22}$$

where $\mathbf{e} = (1, 0, \dots, 0) \in \mathbb{R}^{p+1}$

$$X_{\tilde{\mathbf{x}}} = \begin{pmatrix} 1 & \mathbf{x}_1^{(1)} - \tilde{\mathbf{x}} & \dots & \frac{(\mathbf{x}_1^{(1)} - \tilde{\mathbf{x}})^p}{p!} \\ 1 & \mathbf{x}_1^{(2)} - \tilde{\mathbf{x}} & \dots & \frac{(\mathbf{x}_1^{(2)} - \tilde{\mathbf{x}})^p}{p!} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \mathbf{x}_1^{(n)} - \tilde{\mathbf{x}} & \dots & \frac{(\mathbf{x}_1^{(n)} - \tilde{\mathbf{x}})^p}{p!} \end{pmatrix} \tag{23}$$

[4] FDH stands for free disposable hull.

[5] CCR stands for Charnes, Cooper, and Rhodes who were authors of this model.

and $W_{\tilde{\mathbf{x}}}$ is a diagonal matrix where $w_i(\tilde{\mathbf{x}}) = K((\mathbf{x}_1^{(i)} - \tilde{\mathbf{x}})/h)$ is the weight. Parameter $h > 0$ is a bandwidth parameter and K is a kernel function. We adopted the Gaussian kernel:

$$K(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tag{24}$$

Other applicable kernels, like Epanechnikov kernel, as well as general properties of kernel functions are described e.g. in [4].

In contrast to DEA approaches, the local polynomial regression has two parameters, namely the bandwidth h and p and the quality of fit depends on them. The optimization of p can be done using systematic search since the set finite and small. For each p, the bandwidth h can be optimized using leave-one-approach, details are provided in [4].

Estimating Offset between Two Groups of DMUs. Now, we are about to describe the method itself. We will do it an a step-by-step way:

1. **Benchmark on data** - we use the data D to calculate the $\hat{r}(\mathbf{x}^{(i)})$. We can use one of the benchmarking models provided in the previous paragraphs.
2. **Calculation the residuals** - we introduce residuals as

$$e_i = y_i - \hat{r}(\mathbf{x}^{(i)}) \qquad \forall i = 1, \ldots, n(.)$$

3. **Fitting the residuals** - we calculate the regression model. We adopted the ordinary least square approach with an indicator (dummy) variable as used in [3] that is applicable in the case of model (1).

$$e_i = \beta_0 + \beta_1 \mathbf{1}(k_i = 2) + \xi$$

   where $\mathbf{1}(k_i = 2)$ is indicator that the ith DMU belongs to group 2 and ξ is a zero mean noise.
4. **Point estimate of the offset** - we interpret a regression parameter as a point estimate, namely the regression parameter β_1 can be interpreted as an estimate of Δ because it expresses the average difference between both groups of DMUs.
5. **Estimating the variance of the offset** - we use usual statistical inference on linear regression parameters for variance of the offset, since linear regression [12] estimates β_1 and consequently its variance, too. The estimate will be denoted as b_1 and the variance $SE(b_1)$. It is calculated as follows:

$$SE(b_1) = s\sqrt{\frac{n}{n\sum_{i=1}^{n}(k_i^2) - \left(\sum_{i=1}^{n} k_i\right)^2}}$$

   where

$$s = \sqrt{\frac{\sum_{i=1}^{n}(e_i - \hat{e}_i)^2}{n-2}}.$$

   with $\hat{e}_i$ being output from the regression model in (25)
6. **Construction of interval estimate of the offset** - using this pdf, we can provide the interval estimate as

$$\Delta_{\min} = b_1 + t_{n-2}(\alpha/2) SE(b_1)$$
$$\Delta_{\max} = b_1 + t_{n-2}(1 - \alpha/2) SE(b_1)$$

   for given level of significance $\alpha \in [0, 1]$ where t_{n-2} denotes the cdf of Student distribution with n degrees of freedom.

7. **Density estimation of the offset** - finally, the pdf of Δ is as follows:

$$f(\Delta|D) = f_{n-2}\left(\frac{\Delta - b_1}{SE(b_1)}\right) \tag{25}$$

where f_{n-2} is pdf of Student distribution with $n-2$ degrees of freedom and $SE(b_1)$ is defined in (25).

5 Comparison on Simulated Data

In this Sect. we evaluate the proposed methods on the simulated data so we can evaluate their quality. The data are simulated from models like (2). The evaluation methods use the simulated data only without any knowledge of the models used. We examine general methods for the offset estimation without any prior knowledge and we test if the methods are able to fit the unknown model sufficiently. The use of the simulated data allows us to calculate the offsets analytically from the models and compare them with the data-driven estimates.

5.1 Simulated Data

The following examples are dedicated to the numerical tests and our primary focus is not their real interpretation[6].

Example 1 offers a monotonous, concave function, as expected for the DEA estimation [2] and the noise u_k is left-half-normally distributed.

$$\begin{aligned}
r(x) &= -x^2 + 2x + 15 \\
u_1 &= -|d_1| \qquad d_1 \sim \mathcal{N}(0, \sigma_1) \\
u_2 &= -|d_2| \qquad d_2 \sim \mathcal{N}(0, \sigma_2) \\
\rho &= 1/2 \\
f(\mathbf{x}) &= 1 \qquad \forall \mathbf{x} \in [0,1]
\end{aligned}$$

In this case $\Delta = \delta(\mathbf{x}) = \mathcal{E}[U_2] - \mathcal{E}[U_1] = (\sigma_2 - \sigma_1)\sqrt{2/\pi}$. For the experiments we used $\sigma_1 = 1$ and $\sigma_2 = 2$. The number of instances is $n = 200$.

Example 2 modifies the noise of the previous one so

$$\begin{aligned}
u_1 &\sim \mathcal{N}(\mu_1, \sigma_1) \\
u_2 &\sim \mathcal{N}(\mu_2, \sigma_2)
\end{aligned}$$

Then $\Delta = \mu_2 - \mu_1$. We use $\mu_1 = 1$ and $\mu_2 = 2$ and $\sigma_1 = \sigma_2 = 1$.

Example 3 has same structure as Example 1, but $r(x) = 4x$ for $k = 1$ and $r(x) = 5x$ otherwise. Furthermore, $\sigma_1 = \sigma_2 = 0.2$. From definition (6), it can easily inspected $\Delta = 0.5$ for this case.

[6] Possible interpretation of those models can relate to companies in a segment. The input x can be interpreted as the market share of a company. The output can be the operational costs of the company.

Evaluation Approach We have formulated a set of problems, where the exact values of Δ are analytically known. To compare quality of the proposed estimates, we run the identification procedure $n_\mathrm{s} = 100$ times. Thus, we obtain n_s estimates. We want to calculate how they match the exact values. We measured:

– The quality of the point estimate using MSE $= \frac{1}{n_\mathrm{s}} \sum_{s=1}^{n_\mathrm{s}} (\hat{\Delta}_s - \Delta)^2$ which should be as small as possible,
– The quality of interval estimate using scores whether the exact value is within the interval $[\Delta_{s,\min}, \Delta_{s,\max}]$

$$I = \frac{1}{n_\mathrm{s}} \sum_{s=1}^{n_\mathrm{s}} \mathbf{1}(\hat{\Delta}_{s,\min} \leq \Delta \leq \hat{\Delta}_{s,\max})$$

and it holds that $I \geq 1-\alpha$ is a good result and $I = 1-\alpha$ is a very good result. We used the level of significance $\alpha = 0.05$.
– The probability distribution function as the logarithm of the joint probability $L = \sum_{s=1}^{n_\mathrm{s}} \log f(\Delta|D)$ which should be as big as possible.

Results Tables 1–3 show the results for given tests on the formulated examples. Table 1 shows quite good results in the *MSE* and high *L* for all methods with the exception of *CCR* which fails in all examples[7]. The interval estimates seem not to be very satisfactory since the value should be 0.95 or greater. Only FDH with the value of 0.88 approaches this value. Table 2 shows results for normally distributed noise which are not so good as in the previous case. Table 3 is the only one where CCR was successful. It could be assumed since CCR is dedicated to proportional dependencies between inputs and outputs.

Table 2. Results for Example 1

Approach	*MSE*	*I*	*L*
BCC	0.052	0.640	-5.180
CCR	1.2e8	0.980	-1.2e10
FDH	0.036	0.880	-.3.640
LPR	0.135	0.510	-13.464

5.2 Real Data

We used the same data as in [5] where we assessed the savings achieved by improved control of a heating, air-conditioning and ventilation system. The data set consists of 200 records from an HVAC control system containing (i) daily gas consumption y, (ii) average daily ambient weather x, and index of strategy used during given day k. Since the lower temperature, the higher heating, we used negative values of the ambient temperature. From Fig. 1 it can be seen the CCR leads

[7] The success in the interval estimates for CCR is given by a very wide variance.

Table 3. Results for Example 2

Approach	*MSE*	*I*	*L*
BCC	3.982	0.000	-398.246
CCR	9.0e5	0.960	-9.05e7
FDH	4.142	0.000	-414.182
LPR	3.383	0.000	-338.336

Table 4. Results for Example 3

Approach	*MSE*	*I*	*L*
BCC	1.01	0	-175.74
CCR	0.99	0	-173.93
FDH	1.24	0.32	-4.39
LPR	0.92	0	-199.14

to very flat pdf. Other methods demonstrate the savings, but their estimates of $f(\Delta)$ differ.
From the practical applicability of this evaluation framework, following conclusion can be drawn: if the average achieved savings Δ would be a part of a contract (e.g. the customer pays a ratio of the savings back to the provider), the used evaluation method have to be specified.

6 Conclusions and Future Work

In this text we introduced offset between two groups of DMUs and discussed ways how to estimate it. These estimates have been tested on three examples. We have shown that DEA based estimates are more appropriate for cases where the usual DEA assumptions (one sided noise) are satisfied. The local regression approaches are applicable where those assumptions are not satisfied. Next research shall focus on the estimates of $\delta(\mathbf{x})$ that are of practical importance for evaluation of changes in particular DMUs. Theoretical aspects of the proposed tests shall be subject of deeper investigation since the estimates are not very satisfactory in two of the three examples. Finally, the generalization for multiple-output models shall be addressed.

7 Acknowledgement

This research has been supported by GAČR 13-13502S. The author wishes to express his sincere thanks to Tatiana Valentine Guy and Miroslav Kárný for selfless help with editing and their support in general.

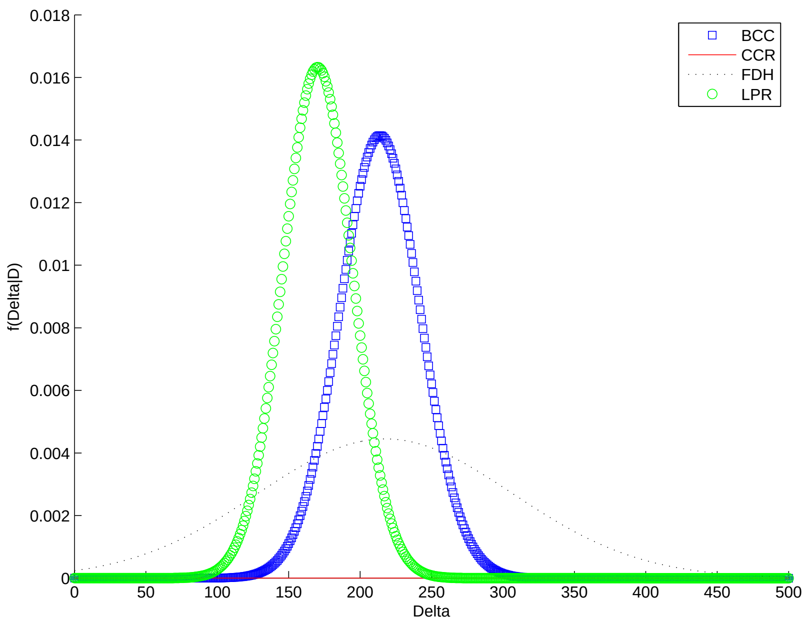

Fig. 1. Pdfs of Δ for particular approaches. BCC and LPR seem to be informative while CCR is practically flat. The achieved savings are very likely between 100 and 300.

References

1. Farrell, M.J.: The Measurement of Productive Efficiency. Journal of the Royal Statistic Society **120**(III) (1957) 253–281
2. Banker, R.: Maximum likelihood, consistency and data envelopment analysis: A statistical foundation. Management Science **39** (1993) 1265–1273
3. Banker, R., Zheng, Z., Natarajan, R.: Dea-based hypothesis tests for comparing two groups of decision making units. European Journal of Operational Research **206**(1) (2010) 231–238
4. Wasserman, L.: All of Nonparametric Statistics. Springer, New York (2006)
5. Macek, K., Mařík, K.: A methodology for quantitative comparison of control solutions and its application to HVAC (heating, ventilation and air conditioning) systems. Energy **44**(1) (2012) 117–125
6. Gomes, E.G., de Mello, J.C.C.B.S., de Freitas, A.C.R.: Efficiency measures for a non-homogeneous group of family farmers. Pesuisa Operacional **32**(3) (2012) 561–574
7. Joo, S.J., Min, H., Kwon, I.W.G., Kwon, H.: Comparative efficiencies of specialty coffee retailers from the perspectives of socially responsible global sourcing. The International Journal of Logistics Management **21** (2010) 490–509
8. Olivares, M., Schenker-Wicki, A.: The dynamics of productivity in the swiss and german university sector: A non-parametric analysis that accounts for heterogeneous production. Technical report, University of Zurich (2012)
9. Banker, R.D., Charnes, A., Cooper, W.W.: Some Models for Estimating Technical and Scale Inefficiencies inData Envelopment Analysis. Management Science **30** (1984) 1078–1092
10. Deprins, D., Simar, L., H.Tulkens: Measuring labor efficiency in post offices. In Marchand, M., Pestieau, P., Tulkens, H., eds.: The Performance of Public Enterprises: Concepts and Measurement. North Holland (1984) 243–257
11. Charnes, A., Cooper, W., Rhodes, E.: Measuring the efficiency of decision making units. European Journal of Operational Research **2**(6) (1978) 429 – 444
12. Kutner, M.H., Nachtsheim, C.J., Neter, J.: Applied Linear Regression Models. Fourth international edn. McGraw-Hill/Irwin (September 2004)