Method for Estimating Confidence Intervals for DEA Efficiency Models Using Regression Models

Filipe Giovani Bonin Bisoffi, Graziella Cardoso Bonadia, Victor Henrique Duarte de Oliveira and Sérgio Ricardo Barbosa

Abstract Benchmarking methods, such as DEA (Data Envelopment Analysis), are used to compare a set of entities regarding their efficiency in a given process. The structure of the DEA method does not take random disturbances into consideration when estimating the efficiency of each entity. In most scenarios, this characteristic does not reflect the reality of the problem, since practically the entire process is subject to external disturbances. Using regression methods, it is possible to generate confidence intervals for DEA-estimated efficiency, considering the model's inputs and outputs as independent variables. With this, the conclusions and subsequent actions based on the returned results are more robust, and begin to contemplate, in a certain manner, random disturbances suffered by the companies.

Keywords DEA · Confidence interval · Nonparametric regression analysis

1 Introduction

Benchmarking is a method used to compare the performance (according to a given metric) of a certain entity, as schools or hospitals, using a reference or benchmark. It is widely used to measure the efficiency of a set of firms or DMUs (Decision-Making Units), as they are generally known in the literature. Given a set of firms with multiple inputs and outputs, a benchmarking method allows to measure the relative performance of the firms through the generation of a curve (frontier) of efficient firms which are obtained by the combination of the providing

F.G.B. Bisoffi (✉) · G.C. Bonadia
Decision Management Support & Applications Department, CPqD - Telecommunications Research and Development Center, Campinas, Brazil
e-mail: fbisoffi@cpqd.com.br

V.H.D. de Oliveira · S.R. Barbosa
CEMIG Distribuição S.A., Belo Horizonte, Brazil
e-mail: srb@cemig.com.br

A. Paradisi et al. (eds.), *Cognitive Technologies*, Telecommunications and Information Technology, DOI 10.1007/978-3-319-53753-5_2

firms. The performance can be measured in terms of the entity productivity or its cost efficiency of inputs to produce outputs.

As observed by Molinero and Woracker [1], benchmarking techniques are especially interesting for organizations that work with inputs or outputs that cannot be monetized. An example of application in this situation can be found in Lorenzett et al. [2], comparing the efficiency of a set of operational units from a network of not-for-profit secondary level professional schools. In this particular example the inputs considered where the total investment and total money transfer of the industry to the school unit. As outputs, the overall performance, the self-sustentation percentual, the client satisfaction index, the product quality index and the quality of management index were considered.

The two main techniques used in benchmarking are the DEA [1] and SFA (Stochastic Frontier Analysis) [3]. The first uses a nonparametric and deterministic structure and the second uses a parametric and stochastic structure. When choosing between a parametric or nonparametric model, one should consider that nonparametric models are more flexible, given that they adapt to the patterns of the data instead of a predefined one. On the other hand, if the chosen model structure is well suited to the underlying nature of the data, the parametric model is expected to give better results. Regarding random effects in the process, stochastic models deal with the errors that are naturally expected in any dataset, either result of measure-related problems or of factors that cannot be measured. One can, thus, test the need to include the error term in the model, and if it is not significant, can opt for a deterministic model.

DEA is based on the solution of linear programming problem to compute the production frontier under which the companies work. This linear problem considers the efficiency as ratio of the outputs produced by the firm and the inputs used in the process. A deficiency of this method is that it does not take random noises into consideration when computing the frontier—consequently, estimates are interpreted deterministically. It is expected that most processes executed by the evaluated companies suffer random disturbances, and not taking this effect into consideration can compromise the efficiency of one or more DMUs assessed in the model.

One way of considering random fluctuations in DEA model results is to estimate confidence intervals for the computed efficiencies. In general, this is accomplished by means of a method known as *bootstrapping*, which consist of resampling the data with replacement to assess the distribution of a given set of estimators. As presented by Bogetoft and Otto [4], there are several alternative approaches to define this procedure. One form of calculating the confidence intervals described by Souza et al. [5], which can be applied to a DEA model with a single product, uses a regression model for this product as a response variable to estimate a confidence interval for the efficiencies. As observed by the author themselves, the literature for multiple product cases is not clear.

The scope of the present study is to build confidence intervals for efficiencies based on regression models trained, considering efficiency a dependent variable, inputs and outputs as independent variables and using DEA-estimated efficiencies as ground-truth to train the regressions. Based on the predicted models, confidence

intervals will be estimated for the efficiency of every DMU and compared with the results of those produced using the bootstrapping method.

2 Methods

Since efficiency can be considered in the interval [0, 1], to estimate confidence intervals it is possible to use the beta regression model proposed by Cribari-Neto and Zeileis [6] or a combination of beta regression models, as described by Grün et al. [7].

Another option is the normal linear regression models. They are adequate for this purpose, as Bogetoft and Otto [4] points out, since they assume that the set of real numbers is the domain of the dependent variable. The Tobit model is an alternative model that presumes normally distributed disturbances, as described in Amemiya [8], and allows variables with lower or upper limits to be modeled in its real domain.

However, these techniques are essentially parametric, in contrast to the DEA method, making it difficult to define an adequate model for the relation between efficiency, inputs and outputs. Therefore, we adopted here nonparametric regression models, for which there is no need to define the relationship between the independent and dependent variables. The estimation methods for this type of regression and other subjects, such as selecting the band matrix, are discussed in Fox [9].

In this work, we built and estimated DEA models using a Variable Return to Scale (VRS) efficiency frontier (also known as BCC model), as defined by Banker, Charnes, and Cooper [10]. We will not present in this paper the linear programming problem associated to DEA, used to compute DMU efficiency. Bogetoft [4] provided the needed theoretical foundation, using the free R Benchmarking package[1] to apply the methodology.

We followed the studies presented by Nadaraya [11] to estimate the nonparametric regressions used to model the DMUs. If ef_i is the calculated efficiency for the i-th DMU and Z_i the vector resulting from the concatenation of inputs X_i and outputs Y_i related to the i-th DMU, then the proposed regression model is described as

$$ef_i = g(Z_i) + \varepsilon_i, \tag{1}$$

where ε_i follows a distribution with $E(\varepsilon_i|Z_i) = 0$ and, thus, $E(ef_i|Z_i) = g(Z_i)$. Unlike parametric regression models, such as the Normal multivariate or beta models, no assumption about $g(Z_i)$ is needed to estimate the model, making this approach more flexible. Normal multivariate regression would assume $g(Z_i) = X\beta$, where X is a matrix—usually nonrandom—containing the observations of the prediction

[1] https://cran.r-project.org/web/packages/Benchmarking/index.html.

variables, and β is a weight vector. Furthermore, it would be assumed that ε_i follows a normal, independent, and homoscedastic distribution. In this case, it is possible to estimate models robust to violation of the residual assumptions, although the relation between the response variable and the predictive variables must be described parametrically.

As a regression model for $g(Z_i)$, we used the nonparametric estimator known as Nadaraya-Watson, which is defined by

$$\widehat{g}_h(z) = \frac{\sum_{j=1}^{n} K_h(z - z_j) ef_j}{\sum_{j=1}^{n} K_h(z - z_j)}, \tag{2}$$

for any kernel $K_h(x)$, with size n of the estimation sample, a nonrandom z vector and the h bandwidth used for estimation.

The kernel function must assume real nonnegative values and be an integrable symmetric function near zero. Furthermore, to ensure the estimated functions are probability distributions, the integral in $\mathbb{R}$ must be equal to one. The kernel proposed by Epanechnikov [12], named after its author, presents the best performance when the evaluated criteria is the average squared error [13]. However, we adopted in the current work the Gaussian kernel, as it is widely used since its estimations are only insignificantly less accurate compared to the Epanechnikov kernel, as proven by Wand and Jones [14], and features convenient mathematical properties.

The Gaussian kernel is defined as

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}, \tag{3}$$

and receives this name because it is based on normal distribution. As for the multivariate case, it is common to accept the product of the marginal kernels in Eq. 2.

The h defines the weight given to each observation for the regression estimate at point z, according to its proximity to z. By choosing a small band, only points closest to z will have any impact on the estimate, while at the other extreme, when using a large band, all observations will have an impact on the estimate. Small values for h are associated with more accurate estimates, while larger values lead to less variable estimates. The band should consider a trade-off between these two equally important characteristics.

This effect can be seen in Fig. 1, which shows the estimate for simulated data density originating from a bimodal normal distribution with an averages −2 and 2. It is clear that the higher band value was unable to properly capture the two true distribution averages, and the lower band value leads to the conclusion that there is more variability in the data distribution than there actually is.

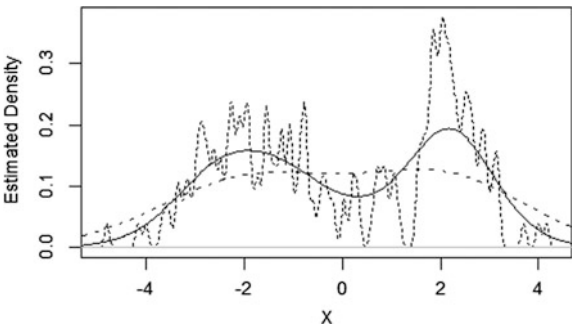

Fig. 1 Comparison of results using different bands. The *dotted line* represents an estimate with a 0.05 band; the *dashed line*, a 1.5 band; and the *solid line*, a 0.6807 band. Notice that a small value for h leads to a overfit on the data and a high value for h leads to a underfit on the data

Ideally, the bands used should, when $n \to \infty$, satisfy $h_n \to 0$ and $nh_n \to \infty$. That is, when the sample size increases, the band must tend to make increasingly local estimates, while increasing the number of observations used to estimate each window.

Several best band selection methods were summarized by Turlach et al. [15], who also present a series of studies on which method is the best. However, their conclusion was that none of the studies reviewed was definitive, and that some of the methods will inevitably have better results in different situations.

The estimates presented in the present work use the band selection method introduced by Li and Racine [16], for generalized multiplicative kernels implemented in the R language inside the NP package,[2] by Hayfield and Racine [17].

The confidence interval for $\widehat{g}_h(z)$ was generated by a bootstrapping procedure. Proposed by Efron [18] based on the *jackknife* method, bootstrapping is a resampling method that generates new samples taken from the original sample, and statistics of interest are computed from these samples. Based on these results, inferences can be made regarding the distribution of these statistics. The concept behind this method is that the bootstrap samples are to the original sample what the original sample is to the population. Since it allows statistical distribution inference without the need for many assumptions, it is a widely used method, especially when the involved algebra becomes too complex, making it hard to identify the theoretical distribution.

3 Data Sets

To evaluate the proposed method, random data sets were created to compute a DEA model and later estimate a nonparametric regression model. Two inputs and four outputs (based on the inputs) were simulated.

[2] https://www.cran.r-project.org/web/packages/np/index.html.

Table 1 Simulation parameter values

	1st round	2nd round
β_1	190	100
β_2	1.5	1.5
β_3	1	1
β_4	90	70
β_5	10	40
β_6	16	50
β_7	0.01	0.1
β_8	15	20

The inputs were simulated as uniform distributions, with the following parametrization:

$$x_1 \sim U[10; 500]$$
$$x_2 \sim U[15; 300]$$

and the outputs with the following equations:

$$y_1 = \beta_1 * \ln(x_1) + \beta_2 * x_2$$
$$y_2 = \beta_3 * x_1 + \beta_4 * \ln(x_2)$$
$$y_3 = \beta_5 * \ln(x_1) + \beta_6 * \ln(x_2)$$
$$y_4 = x_1^{\beta_7} + \beta_8 * \ln(x_2),$$

where β_i, $i = 1, \ldots, 8$ are nonrandom values predefined for each simulation round.

Variables were simulated for 500 DMUs and, to simulate inefficiencies of each DMU, the outputs were multiplied by random reduction factors selected according to a uniform distribution between 0.3 and 1.

The values chosen for β_i were defined according to Table 1.

Once the data sets were simulated, the methods and models described in the previous section were applied to each one of them.

4 Results

DEA models can be defined either as input-oriented, if the main goal is to reduce inputs while keeping outputs producing unaltered, or as output-oriented, if the main goal is to maximize the output while keeping the inputs unaltered an input-oriented DEA model with a VRS frontier was applied to the simulation rounds, without

Table 2 Summary of estimated measured DMU efficiencies

Interval	Quantity	Percentage
0.1 <= ef < 0.2	4	0.8
0.2 <= ef < 0.3	9	1.8
0.3 <= ef < 0.4	23	4.6
0.4 <= ef < 0.5	37	7.4
0.5 <= ef < 0.6	49	9.8
0.6 <= ef < 0.7	59	11.8
0.7 <= ef < 0.8	59	11.8
0.8 <= ef < 0.9	51	10.2
0.9 <= ef < 1	63	12.6
ef = 1	146	29.2

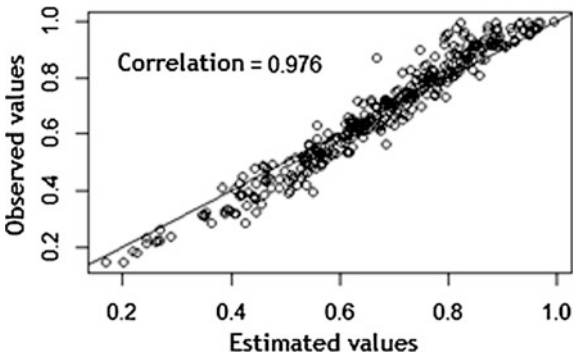

Fig. 2 Relation of N.P. model versus observed values. The *solid line* represents a *straight line* with intercept at the origin and an angular coefficient of one. Ideally, the values should be arranged randomly around the *straight line*. A slight tendency was observed in the lower and higher values, but was not considered detrimental to the model as a whole

slacks,[3] using the R Benchmarking package. Efficiencies for all DMU were computed based on each estimated model.

Table 2 is a summary of the DMU profiles for the first simulation round, as estimated by the DEA model. The quantity and percentage of total DMUs for the entire efficiency interval are also presented. Efficiency values of one were not considered for model estimation. These cases will receive different treatment from the rest.

Figure 2 presents the scatterplot between the estimated values for the efficiency of each DMU using the regression model, and the observed values, computed by the DEA model, together with the correlation between the variables. A strong correlation between the variables can be observed, indicating that the model's accuracy is satisfactory.

[3]Slacks are conditions that can be included to the linear problem to consider output increase in an input-oriented model or input decrease in an output-oriented model.

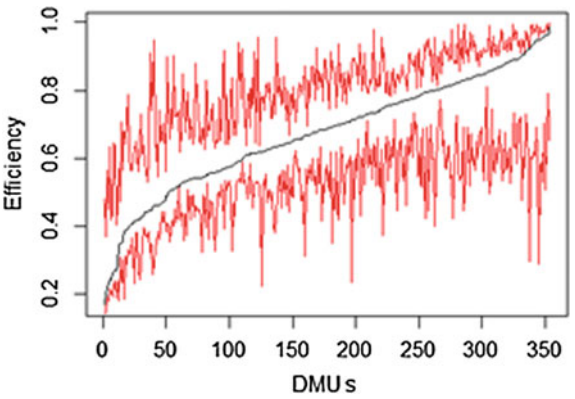

Fig. 3 Ordered efficiency and confidence intervals for the nonparametric model. The *black line* represents the efficiency values estimated by the nonparametric regression model, and the values in *red* represent the upper and lower limits for the confidence interval of 95%

With these results in hand, one hundred bootstrap samples were generated, according to the following algorithm:

1. Repeat 100 times:
   1.1. Sample, with replacement, the observations used for model estimation, with the size of the sample equal to that used to estimate the original model.
   1.2. Estimate the best band for the data set generated in step 1.1.
   1.3. Using the band estimated in step 1.2, compute the estimated efficiency for each point of the original sample.

It was verified that after one hundred rounds, the bootstrap efficiency distribution, estimated by means of the bootstrap sample, did not suffer significant changes, indicating that the number of samples was sufficient for estimator convergence. Figure 3 shows the results achieved by the bootstrap procedure described above. The upper and lower limits were calculated using the 97.5th and 2.5th percentile, respectively, of the generated bootstrap samples.

The idea underlying this procedure is that the result of each DMU is part random and part deterministic, but the efficiency frontier is fixed.

Figure 4 shows the smoothened confidence limits, using the HoltWinters method for time series, without the seasonal component from decomposition, since in the context of the analysis, there is no point in including it. More details on the smoothening method can be checked out in Morettin and Toloi [19]. Smoothening the interval limits should provide more robust results, since part of the variance is absorbed in the process, and make the graphic presentation more readable.

The estimated bias calculated by the bootstrap samples was very close to zero, indicating that the estimates for the generated samples are consistent.

In order to establish a comparison criteria, the bootstrap procedure with bias correction described by Bogetoft and Otto [4] for estimating confidence intervals was also applied to the samples, assuming an input-oriented model with VRS frontier, as in the previous procedure. The R Benchmarking package was used to compute the confidence intervals.

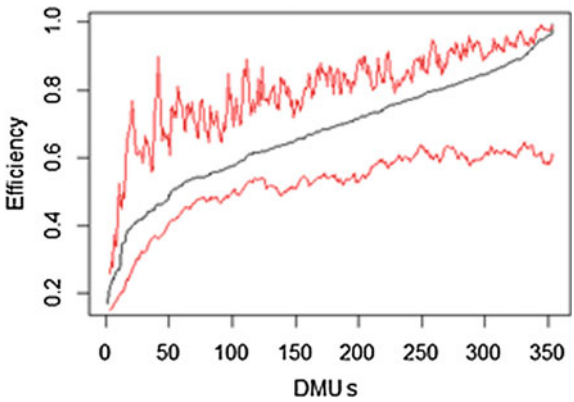

Fig. 4 Ordered efficiency and smoothened confidence intervals for the nonparametric model

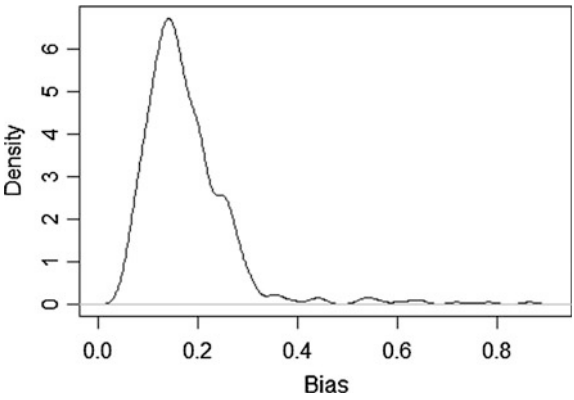

Fig. 5 Estimated density for the bias computed for the first round sample

Figure 5 shows the density estimated in the bootstrap process for the efficiency bias of the DMUs simulated in the first round. Since DEA is positively biased for VRS frontier, in other words, the estimated values for the efficiency of each DMU always overestimate the real value of the parameter, the bias is strictly positive [20]. This happens because the PL built based on the data finds the smallest frontier containing all the observations.

Results of the confidence intervals are displayed in Fig. 6. The black line is the computed efficiency, with no bias correction, for each DMU in the first simulation round, with the values sorted in ascending order.

Efficiencies with bias correction presented incoherent results with negative values and were therefore discarded by the study. The method used in this bootstrap process is based on the concept that computed efficiency is always superior to theoretical efficiency; therefore, the confidence intervals are always lower than the returned values.

As in the previous procedure, HoltWinters smoothening was applied to the confidence interval limits. However, for this case, smoothening was applied only to

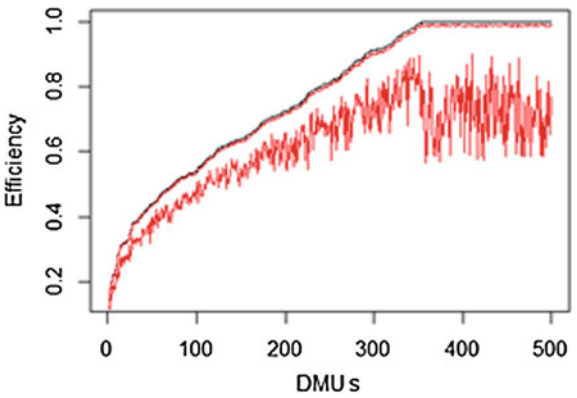

Fig. 6 Ordered efficiency and confidence intervals. The two *red lines* delimit the 95% confidence interval limit for all DMUs

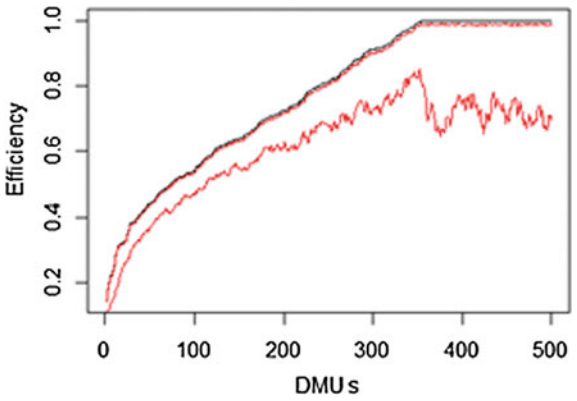

Fig. 7 Ordered efficiency and confidence intervals with smoothened lower limit

the lower limit, since the upper limits presented no significant variance. Figure 7 displays the returned results.

It is interesting to observe that the lower limit of the DMU confidence interval with an efficiency equal to one is lower than the limit of those presenting an efficiency in the [0.7, 1] interval. It would be expected that the limits would follow the value observed for the efficiencies estimated in the original base.

5 Final Considerations

The achieved results show that the two methods studied present conceptual differences. The usual methods, as presented by Bogetoft and Otto [4], are based on the principle that the computed efficiencies are always higher than the real ones, since the frontier computed by the DEA method is the smallest possible. Thus, the objective is to try to define, using a smoothened resampling process, larger frontiers

and generate, for these frontiers, the confidence interval for the efficiencies of each DMU. This way, the information contained in the data can be extrapolated.

The proposed procedure, which uses a nonparametric regression model is based on the idea that each DMU is affected not only by management efficiency, but by random fluctuations as well, and could therefore present better or worse performance than what is displayed. Such random fluctuations are separated from the deterministic part by the nonparametric model, allowing confidence intervals to be generated with the information contained in the data collected originally.

Therefore, it makes no sense to compare the two procedures quantitatively. The choice as to how to build the confidence interval depends on the context of the problem, and the object underlying the analysis. Following the first approach, the intervals are always inferior to the efficiencies originally computed for the DMUs, while the second approach returns intervals that contain the original efficiency.

Furthermore, it is important to point out that the processing time of the first approach is considerably lower than that of the second approach, because of the heavy computational effort required to estimate optimal bands for the nonparametric regression model.

References

1. Molinero CM, Woracker D (1996) Data envelopment analysis. Or Insight 9(4):22–28
2. Lorenzett JR, Lopes ALM, De Lima MVA (2010) Aplicação de método de pesquisa operacional (DEA) na avaliação de desempenho de unidades produtivas para área de educação profissional. Revista Eletrônica de Estratégia & Negócios 3(1):168–190
3. Kumbhakar SC, Lovell CAK (2003) Stochastic frontier analysis. Cambridge University Press
4. Bogetoft P, Otto L (2010) Benchmarking with DEA, SFA, and R. Springer
5. Souza G, Souza M, Gomes E (2011) Computing confidence intervals for output-oriented DEA models: an application to agricultural research in Brazil. J Oper Res Soc 62(10):1844–1850
6. Cribari-Neto F, Zeileis A (2009) Beta regression in R
7. Grün B, Kosmidis I, Zeileis A (2011) Extended beta regression in R: shaken, stirred, mixed, and partitioned. Working papers in economics and statistics
8. Amemiya T (1984) Tobit models: a survey. J Econom 24(1):3–61
9. Fox J (2002) An R and S-plus companion to applied regression. Sage
10. Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manag Sci 30(9):1078–1092
11. Nadaraya EA (1964) On estimating regression. Theor Probab Appl 9(1):141–142
12. Epanechnikov VA (1969) Non-parametric estimation of a multivariate probability density. Theor Probab Appl 14(1):153–158
13. Comaniciu D, Meer P (1999) Mean shift analysis and applications. In: The proceedings of the seventh IEEE international conference on computer vision. IEEE, pp 1197–1203
14. Wand MP, Jones MC (1994) Kernel smoothing. CRC Press
15. Turlach BA et al (1993) Bandwidth selection in kernel density estimation: a review. Université catholique de Louvain
16. Li Q, Racine J (2004) Cross-validated local linear nonparametric regression. Stat Sin 14 (2):485–512

17. Hayfield T, Racine JS (2008) Nonparametric econometrics: the np package. J Stat Softw 27 (5). http://www.jstatsoft.org/v27/i05/
18. Efron B (1979) Bootstrap methods: another look at the jackknife. In: The annals of statistics, pp 1–26
19. Morettin PA, Toloi C (2006) Análise de séries temporais. Blucher
20. Lothgren M (1998) How to bootstrap DEA estimators: a Monte Carlo comparison. WP in Economics and Finance, no 223

http://www.springer.com/978-3-319-53752-8

Cognitive Technologies

Paradisi, A.; Godoy Souza Mello, A.; Lira Figueiredo, F.;
Carvalho Figueiredo, R. (Eds.)

2017, X, 166 p. 64 illus., 41 illus. in color., Hardcover
ISBN: 978-3-319-53752-8