Socio-Economic Planning Sciences 42 (2008) 247–254
SOCIO-ECONOMIC PLANNING SCIENCES
www.elsevier.com/locate/seps

FEAR: A software package for frontier efficiency analysis with *R*

Paul W. Wilson

The John E. Walker Department of Economics, 222 Sirrine Hall, Clemson University, Clemson, South Carolina 29634, USA

Available online 12 March 2007

Abstract

This paper describes a software package for computing non-parametric efficiency estimates, making inference, and testing hypotheses in frontier models. Commands are provided for bootstrapping as well as computation of some new, robust estimators of efficiency, etc.

JEL classification: C14; C15; C44

Keywords: Non-parametric; DEA; Efficiency; Software; Benchmarking

1. Introduction

An extensive literature on measurement of efficiency in production has evolved from the pioneering work of Debreu [1] and Farrell [2]. A large part of this literature uses linear-programming based measures of efficiency along the lines of Charnes et al. [3] and Färe et al. [4]; these methods have been termed as data envelopment analysis (DEA).

Gattoufi et al. [5] cite more than 1800 articles on DEA published in more than 490 refereed journals. DEA and similar non-parametric estimators offer numerous advantages, the most obvious being that one need not specify a (potentially erroneous) functional relationship between production inputs and outputs. Until very recently the non-parametric efficiency literature has ignored statistical issues such as inference, hypothesis testing, etc. However, statistical inference and hypothesis testing are now possible with DEA and other non-parametric efficiency estimators due to results surveyed by Simar and Wilson [6,7].

Standard software packages (e.g., LIMDEP, STATA, TSP) used by econometricians do not include procedures for DEA or other non-parametric efficiency estimators. Several specialized, commercial software packages are available, as well as a small number of non-commercial, free-ware programs; these have been reviewed by Hollingsworth [8–10] and Barr [11]. Academic versions of the commercial packages range in price from about US $280 to US $1500 at current exchange rates.

To varying degrees, the existing packages perform well the tasks they were designed for. Each includes facilities for reading data into the program, in some cases in a variety of formats, and procedures for

E-mail address: pww@clemson.edu.
URL: http://www.clemson.edu/faculty/wilson.

0038-0121/$ - see front matter
doi:10.1016/j.seps.2007.02.001

estimating models that the authors have programmed into their software. A common complaint heard among practitioners, however, runs along the lines of "package *XYZ* will not let me estimate the model I want!" The existing packages are designed for ease of use (again, with varying degrees of success), but at a cost inflexibility, limiting the user to models and procedures the authors have explicitly made available. Moreover, none of the existing packages include procedures for statistical inference. Although the asymptotic distribution of DEA estimators is now known (see [20], for details) for the general case with p inputs and q outputs, bootstrap methods remain the only useful approach for inference. Yet, none of the existing packages include procedures for bootstrapping in frontier models. Similarly, none of the existing packages provide procedures for the newly developed robust alternatives to DEA surveyed in Simar and Wilson [7].

FEAR consists of a software library that can be linked to the general-purpose statistical package *R*. The routines included in *FEAR* allow the user to compute DEA estimates of technical, allocative, and overall efficiency while assuming either variable, non-increasing, or constant returns to scale. The routines are highly flexible, allowing measurement of efficiency of one group of observations relative to a technology defined by a second, reference group of observations. Consequently, the routines can be used to compute estimates of Malmquist indices and their components in any of the decompositions that have been proposed, scale efficiency measures, super-efficiency scores along the lines of Andersen and Petersen [12], and other measures that might be of interest.

Commands are also included to facilitate implementation of the bootstrap methods described by Simar and Wilson [13,14]; in particular, *FEAR*'s `boot.sw98` command can be used to implement the homogeneous bootstrap algorithm described by Simar and Wilson [13]; some programming using commands in *FEAR* and *R* are required when using the current version of *FEAR* to implement the heterogeneous bootstrap algorithm described by Simar and Wilson [14]. Commands in *FEAR* can also be combined to implement statistical tests for irrelevant inputs and outputs or aggregation possibilities as described in Simar and Wilson [15], statistical tests of constant returns to scale versus non-increasing or varying returns to scale as described in Simar and Wilson [16], or the iterated bootstrap procedure described in Simar and Wilson [17].[1] A routine for maximum likelihood estimation of a truncated regression model is included for regressing DEA efficiency estimates on environmental variables as described in Simar and Wilson [21]. In addition, *FEAR* includes commands to estimate Malmquist indices are various decompositions of Malmquist indices, with an implementation of the bootstrap algorithm for described by Simar and Wilson [22]; to compute free-disposal hull (FDH) efficiency estimates [23]; to perform outlier analysis using the methods of Wilson [24]; to compute the new, robust, root-n consistent order-m efficiency estimators described by Cazals et al. [25]; and to compute DEA estimates of cost, revenue, and profit efficiency. A number of these features are unavailable in existing software packages.

2. Where to get *FEAR* and *R*

R is a language and environment for statistical computing graphics. It is an implementation of the *S* language developed at Bell Laboratories, but unlike the commercial version of *S* marketed as *S-Plus* by Lucent Technologies, *R* is freely available under the Free Software Foundation's GNU General Public License. According to the *R* project's web pages (`http://www.r-project.org`),

> *R* provides a wide variety of statistical... and graphical techniques, and is highly extensible... . One of *R*'s strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed... . *R* is an integrated suite of software facilities for data manipulation, calculation and graphical display. It includes (i) an effective data handling and storage facility; (ii) a suite of operators for calculations on arrays, in particular matrices; (iii) a large, coherent, integrated collection of intermediate tools for data analysis; (iv) graphical facilities for data analysis and display either on-screen or on hard copy; and (v) a well-developed, simple and effective programming

[1] The statistical tests described by Simar and Wilson [15,16] do not depend on dubious assumptions of distributions of efficiency scores. Banker [18] and Banker and Natarajan [19] proposed statistics for testing various hypotheses in DEA applications, with critical values chosen either from F-distributions, but the results of Kneip et al. [20] make clear that the statistics proposed by Banker cannot have F distributions. Banker [18] and Banker and Natarajan [19] also proposed Kolmogorov Smirnov tests based on DEA efficiency estimates, but without taking into account the dependence among these estimates discussed by Simar and Wilson [21].

language which includes conditionals, loops, user-defined recursive functions and input and output facilities.

R includes an on-line help facility and extensive documentation in the form of manuals included with the package. In addition, several books describing uses of *R* are available (e.g., [26–28]); see also the recent review by Racine and Hyndman [29]. The current version of *R* can be downloaded from http://lib.stat.cmu.edu/R/CRAN. Pre-compiled binary versions are available for a variety of platforms; at present, however, *FEAR* is being made available only for the Microsoft Windows (XP and later) operating systems running on Intel (and clone) machines.

After downloading and installing *R*, *FEAR* can be downloaded from the following URL:

http://www.economics.clemson.edu/faculty/wilson/Software/FEAR/fear.html

A *Command Reference* and *User's Guide* can also be downloaded from this site; the *User's Guide* gives instructions for installing *FEAR* into *R*, as well as a number of examples illustrating some of the capabilities of *FEAR*. The web site also contains licensing information; while use of *FEAR* is free for academic purposes, potential government, commercial, or other users outside an educational institution should contact the author

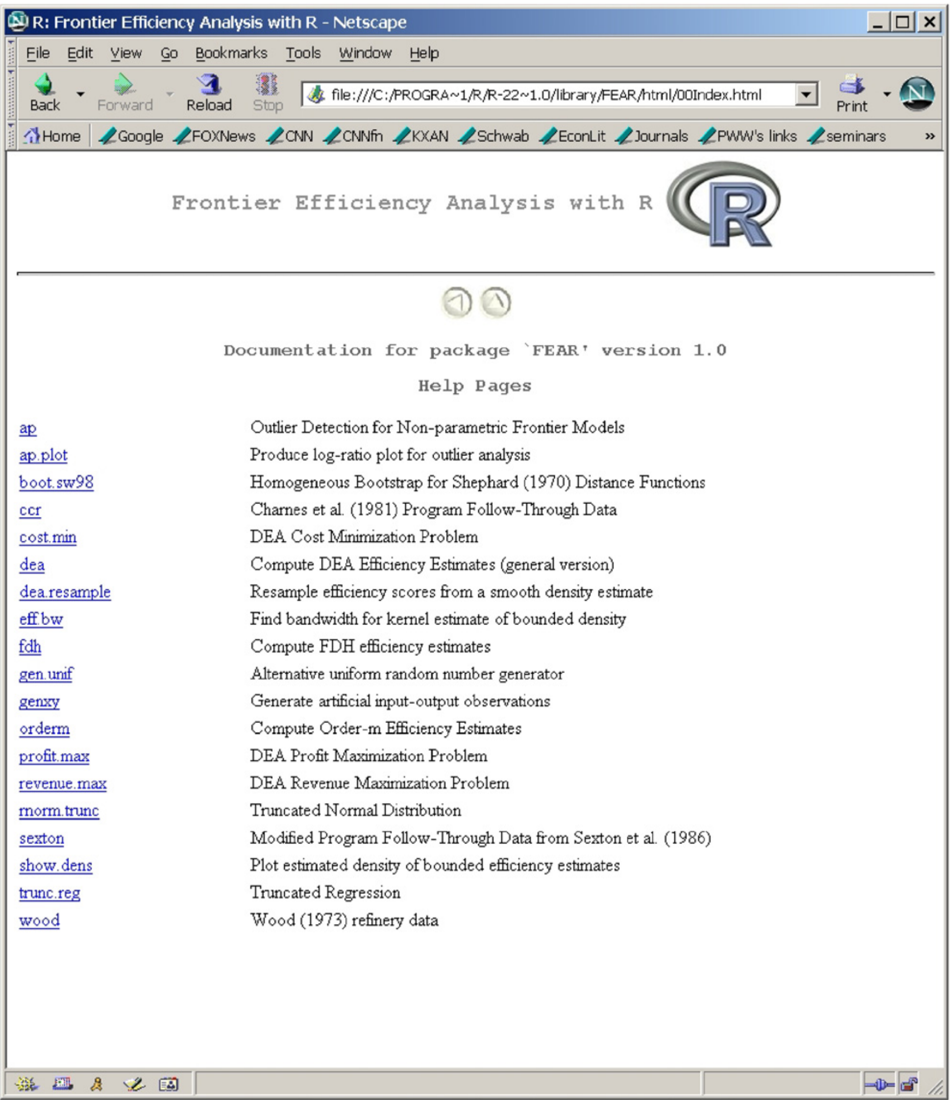

Fig. 1. List of commands in the FEAR package.

before using the *FEAR* software. The current version of *FEAR* (at the time of this writing) is 1.0; as features are added and modifications are made, new versions will be released (with higher version numbers). Every attempt will be made to maintain backward-compatibility with earlier versions of *FEAR*, and versions prior to the current release will remain available for some time.

3. Using *FEAR* and *R*

Once *R* and *FEAR* have been downloaded and installed, *R*'s graphical user interface (GUI) can be started by clicking on the desktop *R* icon; commands are typed at a prompt in the "RConsole" window contained within the RGui window. After installing *FEAR*, *R*'s on-line help facility can be used to find documentation on the commands implemented in *FEAR*, as shown in Fig. 1. Clicking on a command name displays a page giving details on use of the command, including arguments that must be passed, optional arguments and any defaults, etc., as well as a detailed description of what is returned by the command.

FEAR includes data that can be used to illustrate the library's capabilities. First, the command `library(FEAR)` must be typed in the RConsole window to load the FEAR library, as shown in Fig. 2. One might then type `help(ccr)` to learn that the Charnes et al. [30] data contain 5 inputs used to produce 3 outputs among 70 schools. The following commands load these data and then organize the input vectors in a (5×70) matrix and the output vectors in a (3×70) matrix:

```
data(ccr)
x = t(matrix(c(ccr$x1, ccr$x2, ccr$x3, ccr$x4, ccr$x5),
      nrow = 70,ncol = 5))
y = t(matrix(c(ccr$y1,ccr$y2,ccr$y3),nrow = 70,ncol = 5))
```

The following commands reproduce the outlier analysis in Wilson [24]:

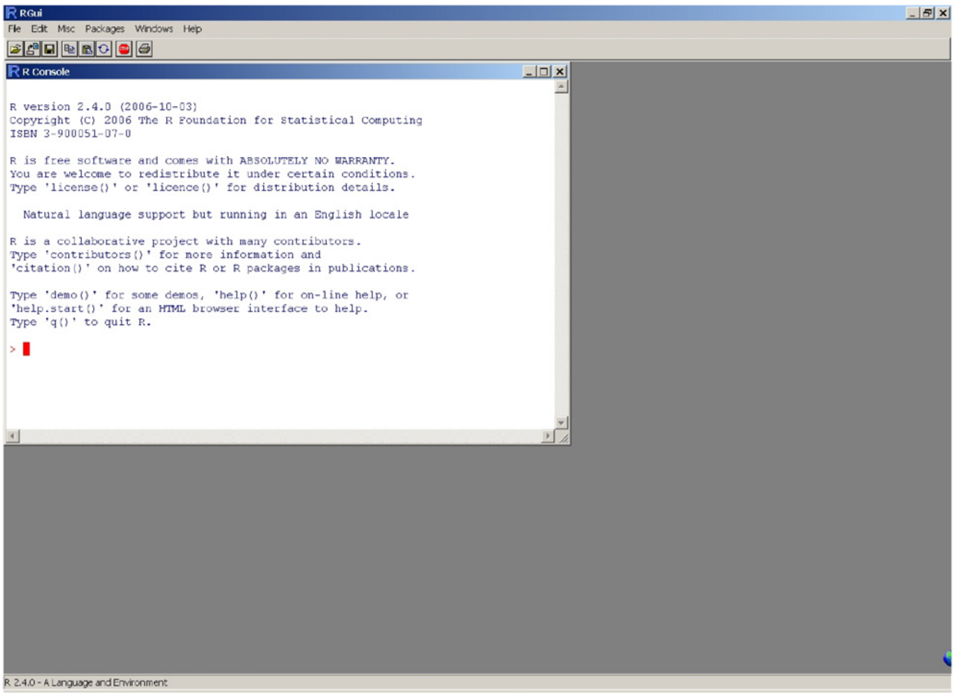

Fig. 2. The *R* graphical user interface.

```
tmp = ap(x,y,NDEL = 12)
windows()
ap.plot(RATIO = tmp$ratio)
```

The resulting plot is shown in Fig. 3.

Next, Shephard [31] output distance function estimates can be computed:

```
d.out = dea(x,y,ORIENTATION = 2)
```

R's `summary` command

```
summary(d.out)
```

gives summary statistics on the estimates. Alternatively, the `summary` command can be combined with the `ifelse` statement to obtain summary statistics on the distance function estimates that are different from unity:

```
summary(ifelse(d.out = = 1,NA,d.out))
```

as shown in Fig. 4. *FEAR*'s `show.dens` command can be used to produce plots of kernel density estimates for the distance function estimates contained in `d.in` and `d.out`; the necessary bandwidth parameters can be obtained with *FEAR*'s `eff.bw` command:

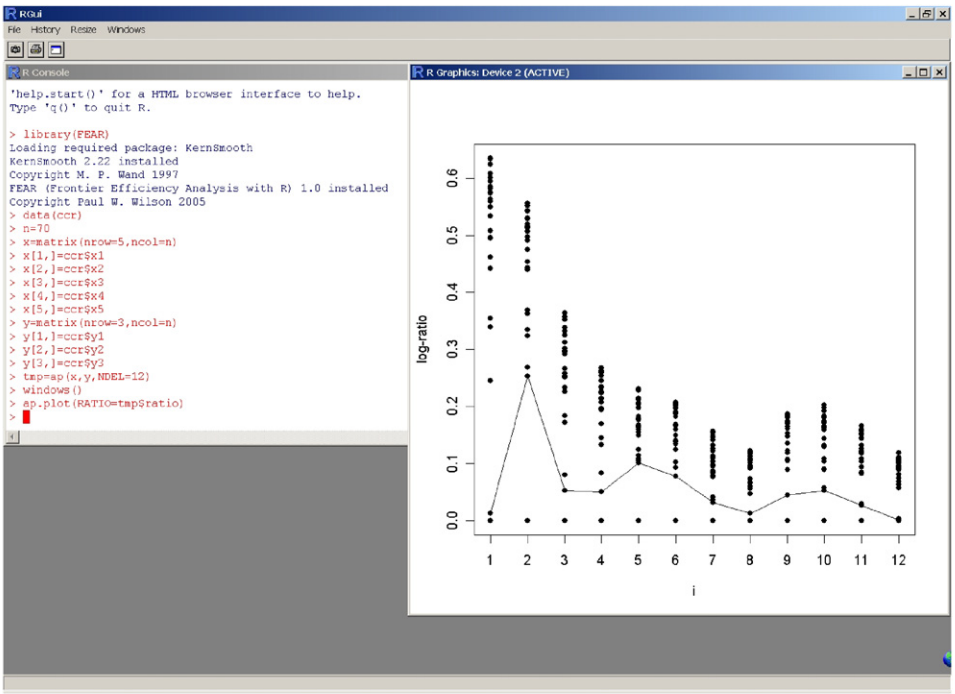

Fig. 3. Plots produced by the `ap.plot` command.

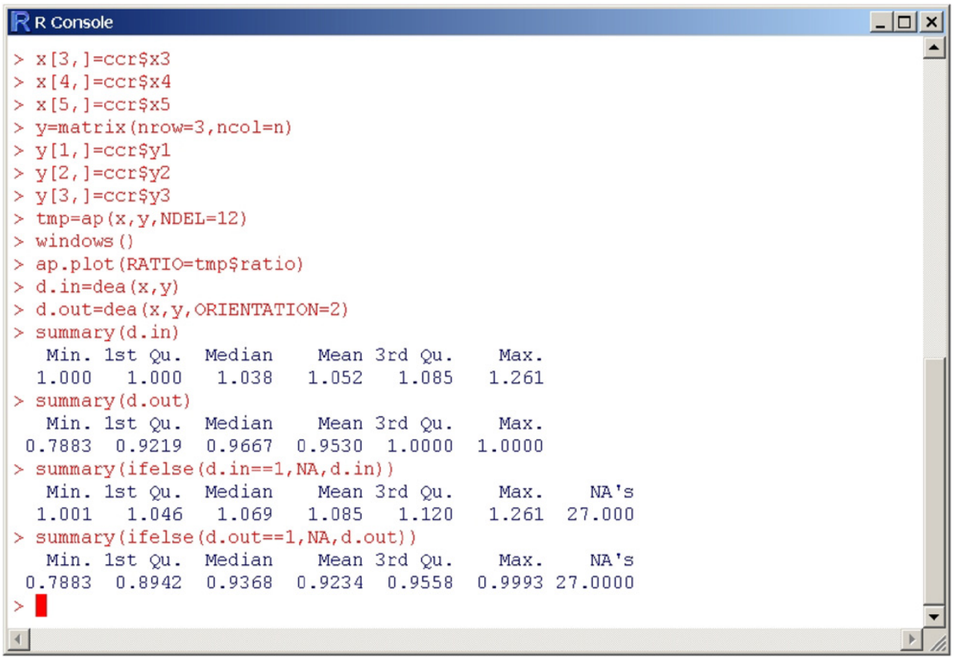

Fig. 4. Output from *R*'s `summary` command.

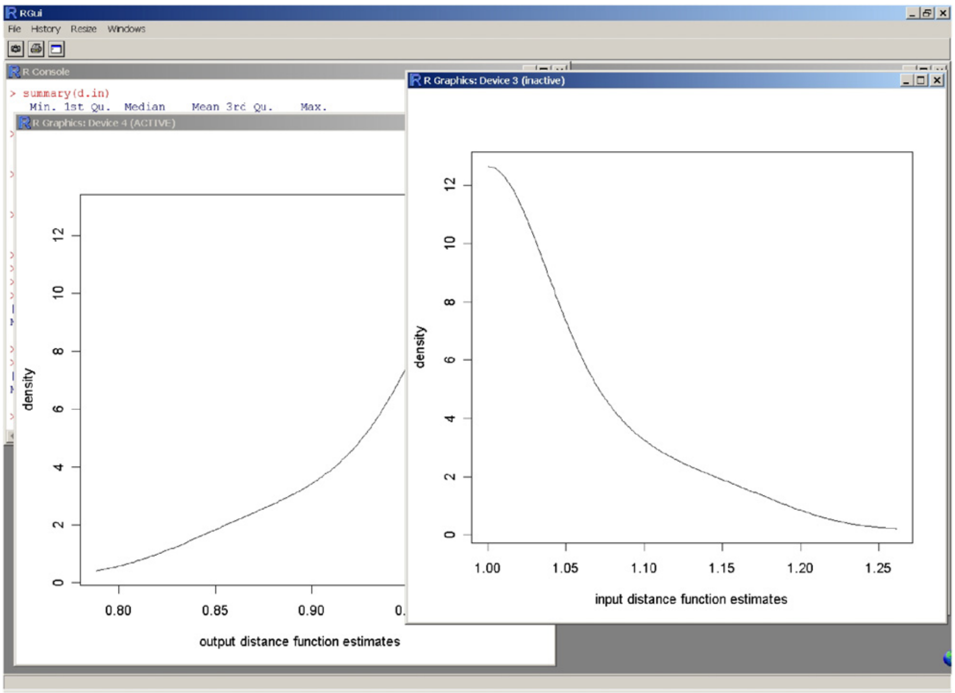

Fig. 5. Kernel density plots produced by the `show.dens` command.

```
h = eff.bw(d.out)
show.dens(dist = d.out,bw = h,
      XLAB = "output distance function estimates")
```

The resulting plots are shown in Fig. 5; here, they have been drawn on the user's screen, but alternatively, *R*'s `postscript` command can be used to write postscript code for the plots to a file for inclusion in a manuscript, etc. The `show.dens` command uses a reflection method to avoid inconsistency problems for the kernel density estimator and the lower (upper) boundary 1 for the input (output) efficiency estimates.

The command `boot.sw98` implements the bootstrap algorithm described by Simar and Wilson [13]; this command can be used to estimate 95% confidence intervals for the input distance functions corresponding to each observation in the Charnes et al. [30] data by typing

```
result = boot.sw98(XOBS = x,YOBS = y,DHAT = d.out,ORIENTATION = 2)
```

The following command will produce LATEX code for a table displaying results from the last command; the first column contains the observation number, the second column contains the input distance function estimates, while the third and fourth columns contain the lower and upper bounds of the estimated confidence intervals:

```
tmp = paste(c(1:70), " &",d.in," &",
      result$conf.int[,1],"&",
      result$conf.int[,2],"\\", sep = " ")
```

4. Concluding remarks

FEAR is a very flexible, extensible package unlike any currently available for estimation of productivity and efficiency. The cost of this flexibility is that the user must type commands at a command-line prompt; for some, this may be unsatisfactory. For others, however, *FEAR* will be a useful tool, allowing one to estimate quantities that have not been explicitly programmed into other packages, and that have perhaps not been anticipated by this or other authors.

In addition to the commands illustrated above, the current version *FEAR* includes commands to compute DEA estimates of cost, revenue, and profit efficiency, Malmquist indices and various decompositions, and order-*m* efficiency estimates.

References

[1] Debreu G. The coefficient of resource utilization. Econometrica 1951;19:273–92.
[2] Farrell M. The measurement of productive efficiency. Journal of the Royal Statistical Society A 1957;120:253–81.
[3] Charnes A, Cooper W, Rhodes E. Measuring the inefficiency of decision making units. European Journal of Operational Research 1978;2:429–44.
[4] Färe R, Grosskopf S, Lovell C. The measurement of efficiency of production. Boston: Kluwer-Nijhoff Publishing; 1985.
[5] Gattoufi S, Oral M, Reisman A. Data envelopment analysis literature: a bibliography update (1951–2001). Socio-Economic Planning Sciences 2004;38:159–229.
[6] Simar L, Wilson P. Statistical inference in nonparametric frontier models: the state of the art. Journal of Productivity Analysis 2000;13:49–78.
[7] Simar L, Wilson P. Statistical inference in nonparametric frontier models: recent developments and perspectives. In: Fried H, Lovell C, Schmidt S, editors. The measurement of productive efficiency. 2nd ed. Oxford: Oxford University Press; 2007 [chapter 4], in press.
[8] Hollingsworth B. A review of data envelopment analysis software. Economic Journal 1997;107:1268–70.
[9] Hollingsworth B. Data envelopment analysis and productivity analysis: a review of the options. Economic Journal 1999;109:F458–62.
[10] Hollingsworth B. Non parametric efficiency measurement. Economic Journal 2004;114:F307–11.

[11] Barr R. DEA software tools and technology. In: Cooper W, Seiford L, Zhu J, editors. Handbook on data envelopment analysis. Boston: Kluwer Academic Publishers; 2004. p. 539–66.
[12] Andersen P, Petersen N. A procedure for ranking efficient units in data envelopment analysis. Management Science 1993;39:1261–4.
[13] Simar L, Wilson P. Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. Management Science 1998;44:49–61.
[14] Simar L, Wilson P. A general methodology for bootstrapping in non-parametric frontier models. Journal of Applied Statistics 2000;27:779–802.
[15] Simar L, Wilson P. Testing restrictions in nonparametric efficiency models. Communications in Statistics 2001;30:159–84.
[16] Simar L, Wilson P. Nonparametric tests of returns to scale. European Journal of Operational Research 2001;139:115–32.
[17] Simar L, Wilson P. Performance of the bootstrap for DEA estimators and iterating the principle. In: Cooper W, Seiford L, Zhu J, editors. Handbook on data envelopment analysis. New York: Kluwer Academic Publishers, Inc.; 2004. p. 265–98 [Chapter 10].
[18] Banker R. Hypothesis tests using data envelopment analysis. Journal of Productivity Analysis 1996;7:139–59.
[19] Banker R, Natarajan R. Statistical tests based on dea efficiency scores. In: Cooper W, Seiford L, Zhu J, editors. Handbook on data envelopment analysis. New York: Kluwer Academic Publishers, Inc.; 2004. p. 265–98 (Chapter 11).
[20] Kneip A, Simar L, Wilson P. Asymptotics for DEA estimators in nonparametric frontier models. Discussion Paper #0317, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium, 2003.
[21] Simar L, Wilson P. Estimation and inference in two-stage. semi-parametric models of productive efficiency, Journal of Econometrics 2007;136:31–64.
[22] Simar L, Wilson P. Estimating and bootstrapping Malmquist indices. European Journal of Operational Research 1999;115:459–71.
[23] Deprins D, Simar L, Tulkens H. Measuring labor inefficiency in post offices. In: Pestieau MMP, Tulkens H, editors. The performance of public enterprises. Amsterdam: North-Holland; 1984. p. 243–67.
[24] Wilson P. Detecting outliers in deterministic nonparametric frontier models with multiple outputs. Journal of Business and Economic Statistics 1993;11:319–23.
[25] Cazals C, Florens J, Simar L. Nonparametric frontier estimation: a robust approach. Journal of Econometrics 2002;106:1–25.
[26] Dalgaard P. Introductory statistics with R. New York: Springer; 2002.
[27] Venables W, Ripley B. Modern applied statistics with S. New York: Springer; 2002.
[28] Verzani J. Using R for introductory statistics. London: Chapman & Hall; 2004.
[29] Racine J, Hyndman R. Using R to teach econometrics. Journal of Applied Econometrics 2002;17:175–89.
[30] Charnes A, Cooper W, Rhodes E. Evaluating program and managerial efficiency: an application of data envelopment analysis to program follow through. Management Science 1981;27:668–97.
[31] Shephard R. Theory of cost and production functions. Princeton, NJ: Princeton University Press; 1970.