orderalpha: non-parametric order-α Efficiency Analysis for Stata

Harald Tauchmann

Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI)

July 1st, 2011

2011 German Stata Users Group Meeting, University of Bamberg

Introduction

- Countless empirical efficiency analyses in economics
  - **D**ecision **M**aking **U**nits/production units (**DMU**s): firms, states, universities, etc.
- Measuring distance to *Production Possibility Frontier*
  - alternatively: cost- or profit-frontier
- *Output*-orientated and *Input*-orientated efficiency
  1. How much additional output $(y_1, \ldots, y_L)$ can be produced leaving inputs $(x_1, \ldots, x_M)$ unchanged?
  2. How much input $(x_1, \ldots, x_M)$ can be saved leaving outputs $(y_1, \ldots, y_L)$ unchanged?
- Prime purpose: estimating efficiency scores θ_i for individual DMUs
  - (Input-oriented) efficiency score $(0, 1]$: factor by which input consumption can *proportionally* be reduced

Parametric Approaches: Stochastic Frontier Models

- Introduced by Aigner et al. (1977)

$\rightarrow$ stata command **`frontier`**

- Production- (cost-, profit-) function estimated
- Linear regression model augmented by additional non-positive (reps. non-negative) error term η_i
- Maximum-likelihood estimation
- Distributional assumptions for η_i and υ_i required

$$\log(y_i) = \beta_0 + \sum_{m=1}^{M} \beta_M x_{Mi} + \upsilon_i - \eta_i \quad \text{with} \quad \eta_i \geq 0$$

- (Output-oriented) technical efficiency score computed as: $\hat{\theta}_i^{SF} = \mathrm{E}(\exp\left(-\eta_i\right) | \upsilon_i - \eta_i)$

Non-Parametric Approaches

1. Data envelopment Analysis (DEA)

- Introduced by Charnes et al. (1978)

→ stata ado-file **dea** (Ji & Lee, 2010)

- Linear programming approach
- Envelopes data by *piecewise-linear convex* hull
- Solution for θ (input-oriented) efficiency score $\hat{\theta}_i^{DEA}$:

$$\min_{\theta,\lambda} \theta \qquad \text{subject to}$$

$$\begin{aligned}
\theta x_{mi} - \sum_{j=1}^{N} \lambda_j x_{mj} &\geq 0 \quad m = 1, \ldots, M \\
\sum_{j=1}^{N} \lambda_j y_{lj} - y_{li} &\geq 0 \quad l = 1, \ldots, L \\
\lambda_j &\geq 0 \quad \forall j
\end{aligned}$$

Non-Parametric Approaches II

2. Free Disposal Hull (FDH)

- Introduced by Deprins et al. (1984)
- Based on principle of weak dominance
  - DMU *i* compared to those DMUs that produce at least the same amount of any output
  - DMU with minimal input use serves as reference
- Envelopes data by piecewise linear *non-convex* (step) hull
- (Input-oriented) efficiency scores computed as:

$$\hat{\theta}_i^{FDH} = \min_{j=1,\ldots,N \mid y_{lj} \geq y_{li} \, \forall \, l} \left\{ \max_{m=1,\ldots,M} \left\{ \frac{x_{mj}}{x_{mi}} \right\} \right\}$$

Parametric vs. Non-Parametric Approaches

1. Parametric Approach (shortcomings):
   - Relies on distributional assumptions
   - Functional form for production technology required
   - Production function ill-suited regression model (endogeneity of inputs)
   - Accommodates only single-output technologies

2. Non-Parametric Approach (shortcomings):
   - No well-defined data generating process ($\rightarrow$ recent advances)
   - Deterministic approach

   $\Rightarrow$ Extremely vulnerable to outliers and measurement error

Partial Frontier Approaches

- Sensitivity to outliers reduced by allowing for super-efficient DMUs
- Super-efficient DMUs located beyond production-possibility frontier
- partial frontier envelopes just a sub-sample of the data
- Super-efficiency: (input-oriented) efficiency score > 1

Partial frontier approaches that generalize FDH:

1. Order-*m* efficiency (Cazals et al., 2002)
2. Order-α efficiency (Aragon et al., 2005)

Order-*m* Efficiency

- Adds a 'layer of randomness' to FDH
- Series of FDH analyses using an randomly drawn sup-samples of size m
- DMU may or may not serve as its own peer
- Re-sampling repeated D times
- $\hat{\theta}_i^{OM}$: average of D efficiency sores
- Shortcoming: very time consuming (for large data sets)
  - $\rightarrow$ Rules virtually out statistical inference based on bootstrapping
  - $\rightarrow$ Determining appropriate value for m may require trying numerous values

Order-α Efficiency

- Chooses $[100-\alpha]^{th}$ (α^{th}) percentile, with $0 \leq \alpha \leq 100$, rather an minimum (maximum) as efficiency benchmark
- Not **min**. input consumption, but $[100-\alpha]^{th}$ **percentile** (among peer-DMUs) serves as reference
- FDH represent special case of order-α (for $\alpha = 100$)
- No re-sampling $\rightarrow$ less time consuming than order-m

$\rightarrow$ New stata ado-file **`orderalpha`**

Order-α Input-Oriented Efficiency:

$$\hat{\theta}_i^{OA_{input}} = \underset{j=1,\dots,N|y_{lj} \geq y_{li}\,\forall\, l}{P_{100-\alpha}} \left\{ \max_{m=1,\dots,M} \left\{ \frac{x_{mj}}{x_{mi}} \right\} \right\}$$

Order-α Output-Oriented Efficiency:

$$\hat{\theta}_i^{OA_{output}} = \underset{j=1,\dots,N|x_{mj} \leq x_{mi}\,\forall\, m}{P_{\alpha}} \left\{ \min_{l=1,\dots,L} \left\{ \frac{y_{lj}}{y_{li}} \right\} \right\}$$

Illustration for Single-Input Single-Output case

(districts-level health production in Germany)

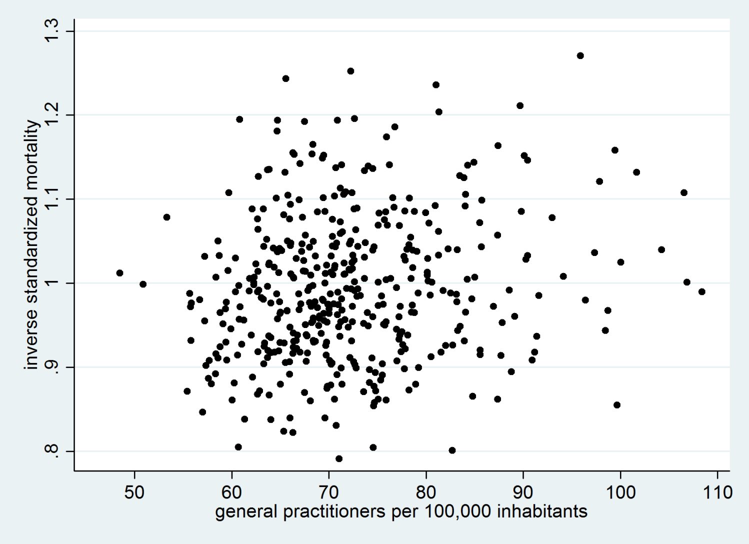


Illustration for Single-Input Single-Output case

(districts-level health production in Germany)

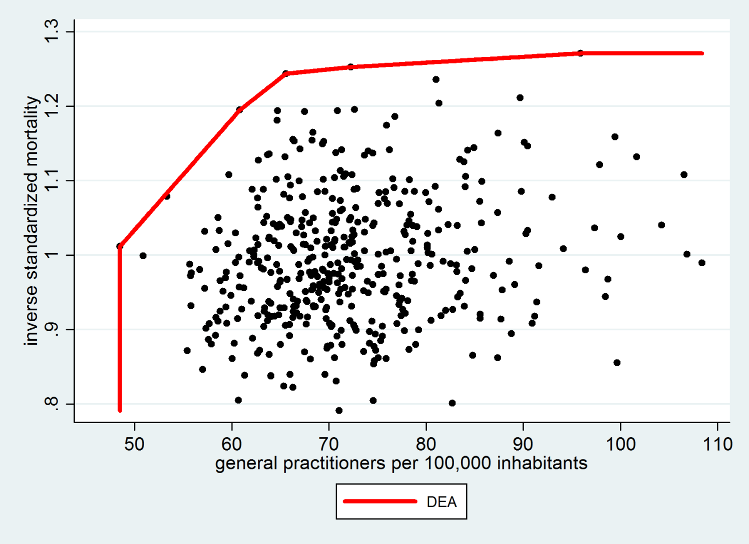


Illustration for Single-Input Single-Output case

(districts-level health production in Germany)

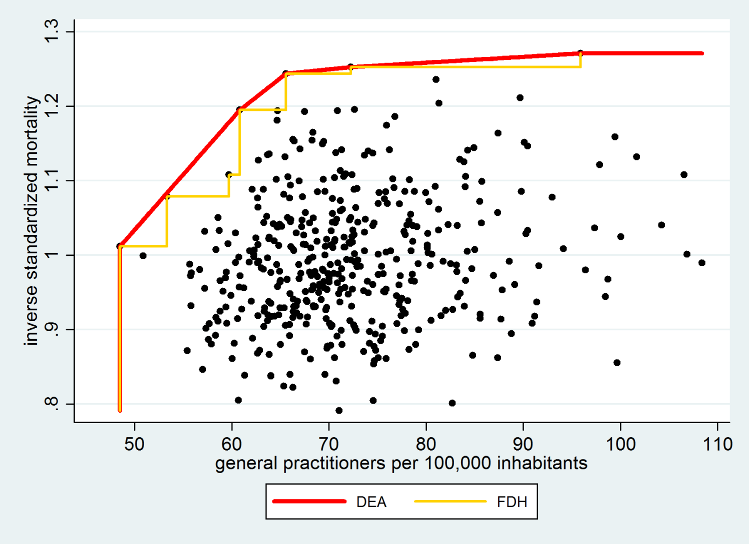


Illustration for Single-Input Single-Output case

(districts-level health production in Germany)

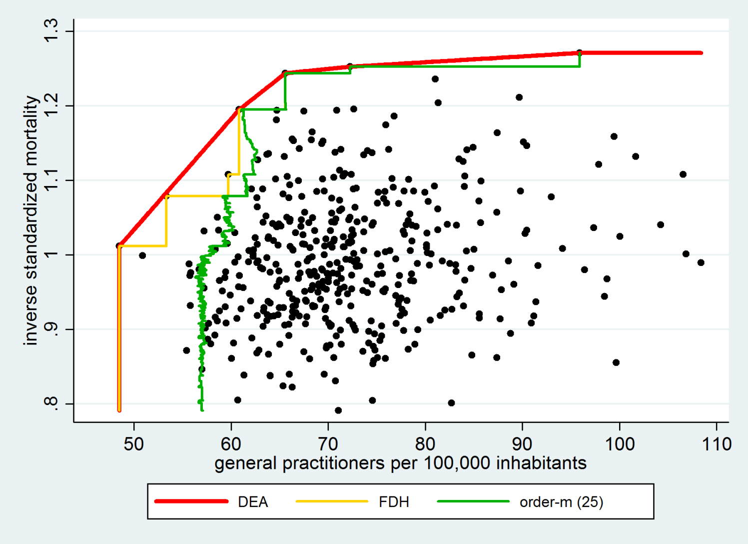


Illustration for Single-Input Single-Output case

(districts-level health production in Germany)

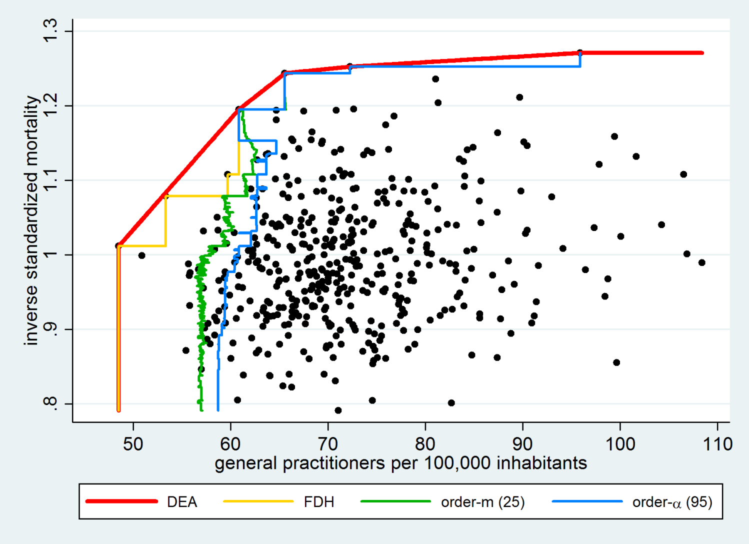
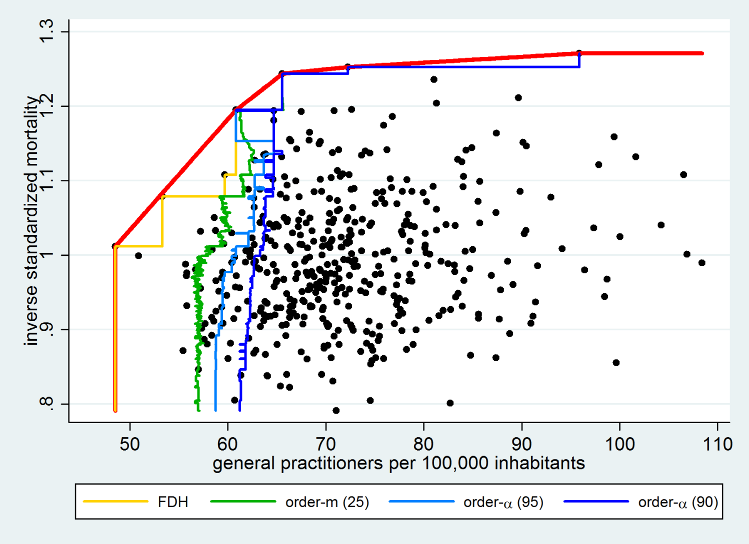


Illustration for Single-Input Single-Output case

(districts-level health production in Germany)



orderalpha: Syntax

Viewer (#5) [help orderalpha]

help orderalpha

Advice | Contents | What's New | News

help orderalpha

Title

orderalpha — Order-alpha efficiency analysis

Syntax

orderalpha *varname* [*if*] [*in*], **inputs**(*varlist1*) **outputs**(*varlist2*) [*options*]

technology_definition	description
Model	
varname	identifier
inputs(*varlist1*)	list of input variables
outputs(*varlist2*)	list of output variables

options	description
Main	
ort(*string*)	*string* may be **output** or **input**; default is **ort(***input***)**
alpha(*#*)	set benchmark percentile; default is **alpha(***100***)**
SE/Bootstrap	
bootstrap	perform bootstrap using 100 replications
reps(*#*)	set (temporary) number *#* of bootstrap replications and perform bootstrap
tune(*#*)	set tuning parameter for subsampling bootstrap; values within the [0.5,1] interval are allowed
Reporting	
level(*#*)	set confidence level; default is **level(***95***)**
table(*string*)	display table of results; *string* may be **scores** or **full**, **no** table is the default
dots(*#*)	display replication/loop dots; *#* may be **1** or **2**, no dots is the default
invert	report reciprocal of output-oriented efficiency scores
Generate	
generate(*newvarlist*)	supply names for new variables, containing efficiency scores, ranks, and reference dmus
replace	replace existing variables in *newvarlist*
nogenerate	do not create new variables containing results

weights are not allowed; see weight.
bootstrap, **by**, and **svy** are not allowed; see prefix.

Description

Application: Regional Health Production in Bavaria

- Decision making units:
  - Districts ('kreis') in Bavaria
  - Cross section: year 2004
  - # of observations: 96
- Inputs:
  i. Resident medical specialists per 100 000 inhabitants ('specialists')
  ii. General practitioners per 100 000 inhabitants ('gps')
  iii. Hospital beds per 10 000 inhabitants ('beds')
- Outputs:
  i. Inverse normalized (to district demographics) *mortality* ('survival')

orderalpha: Screen Shot (regional health production in Bavaria)

```
. orderalpha kreis if jahr == 2004 & regio >= 9000 & regio < 10000, inputs(specialists
> gps beds) outputs(survival) alpha(95) replace table(full) reps(200) gen(escore erank
> refdmu)

Order-alpha(95) input-oriented efficiency scores estimated (variable escore)

Number of dmus                 = 96
Number of inputs               = 3
Number of outputs              = 1
Mean efficiency                = .9229
Median efficiency              = .9023
Share of efficient dmus        = .2083
Share of super-efficient dmus  = .0833

-----------------------------------------------------------------------------
    dmu (kreis) |  Eff. Score    Std. Err.      z Stat.    Eff. Rank    Ref. DMU
----------------+------------------------------------------------------------
Aichach-Friedbe |          1     .1322505            0            9  Aichach-Fr
Altötting, Land |   .8259127     .0254798     6.832376           65  Eichstätt,
Amberg, krsfr.  |    .650349     .0157609     22.18474           89  Aschaffenb
Amberg-Sulzbach |   1.104511     .3570595     .2926985            5  Landshut,
Ansbach, Landkr |          1     .1559544            0            9  Ansbach, L
Ansbach, krsfr. |   .7476639     .0246253     10.24702           77  Freising,
Aschaffenburg,  |   2.423852     .7841469     1.815798            2  Aichach-Fr
Aschaffenburg,  |   .8616958      .035989     3.842959           57  Aschaffenb
Augsburg, Landk |   .9280609     .1040334     .6914996           43  Regensburg
Augsburg, krsfr |   .7404743     .0212208      12.2298           79  Landshut,
Bad Kissingen,  |   .8131074     .0247055     7.564818           70  Straubing-
Bad Tölz-Wolfra |          1     .5837663            0            9  Bad Tölz-W
Bamberg, Landkr |   1.082539     .3349955     .2463872            6  Coburg, La
Bamberg, krsfr. |   .5726839     .0222016     19.24707           96  Erlangen-H
Bayreuth, Landk |   1.435832     .5069627     .8596922            3  Bamberg, L
Bayreuth, krsfr |   .7540368     .0182821     13.45376           76  Aschaffenb
Berchtesgadener |    .787752      .056467       3.7588           73  Erding, La
```

orderalpha: Saved Results

Viewer (#5) [help orderalpha]

help orderalpha

Advice | Contents | What's New | News

Saved results

orderalpha saves the following in **e()**:

Scalars

e(N)	number of dmus
e(alpha)	value of **alpha(#)**
e(inputs)	number of inputs
e(outputs)	number of outputs
e(efficient)	share of efficient dmus
e(super)	share of super-efficient dmus
e(mean_e)	mean estimated efficiency
e(med_e)	median estimated efficiency
e(level)	confidence level

Macros

e(cmd)	**orderalpha**
e(cmdline)	command as typed
e(title)	**Order-alpha efficiency analysis**
e(dmuid)	*varname* (name of dmu-identifier)
e(model)	either **Order-alpha** or **FDH**
e(saved)	names of variables saved (not saved for option **nogenerate**)
e(table)	**scores**, **full**, or **no**
e(invert)	either **inverted** or **notinverted** (not saved for **ort(***input***)**)
e(ort)	either **input** or **output**
e(properties)	either **b** or **b V**
e(depvar)	**dmu**

Matrices

e(b)	vector of estimated efficiency scores (*colnames* are of the form *varname:value_of_varname*)
e(ranks)	vector of efficiency ranks (*colnames* are of the form *varname:value_of_varname*)
e(reference)	matrix of names of reference dums (not saved if *varname* is a string variable)

Functions

e(sample)	marks estimation sample

orderalpha, boot reps(#) additionally saves the following in **e()**:

Scalars

e(N_reps)	number of bootstrap repetitions
e(tune)	value of tuning parameter
e(N_bs)	size of bootstrap samples

Macros

e(vce)	**bootstrap**
e(vcetype)	**Bootstrap**

Matrices

e(V)	bootstrap variance-covariance matrix for estimated efficiency scores
e(bias)	estimated biases
e(reps)	number of nonmissing results
e(b_bs)	bootstrap estimates

Order-α based outlier-detection

Idee proposed by Daraio & Simar (2007):

- Increasing the value of α reduces number of DMUs classified as “super-efficient”
- In the absence of outliers: share of super-efficient DMUs should decrease smoothly
- Discontinuity points at presence of outliers

$\Rightarrow$ DMUs still classified “super-efficient” for $\alpha \geq \alpha^{disc}$ (point of discontinuity) most likely outliers

$\Rightarrow$ To be excluded from efficiency analysis applying FDH or DEA

Implementing approach of Daraio & Simar (2007)

`oaoutlier`:

- Carries out series of order-α analyses
- Plots share of super-efficient units against α
- Suggests two local and one global rules for detecting discontinuities:
  1. α for which the twice differenced series takes minimum value (following a non-negative one)
  2. Values of α for which negative values persist after repeatedly smoothing twice differenced series by running odd-spaced median smoothers ($\rightarrow$ **`smooth`**)
  3. α that minimizes BIC for splitting the series into two parts and fitting linear (quadratic) functions to each

oaoutlier: Syntax

Viewer (#6) [help oaoutlier]

help oaoutlier

Advice | Contents | What's New | News

help oaoutlier

Title

oaoutlier — Outlier detection based on order-alpha efficiency analysis

Syntax

oaoutlier *varname* [*if*] [*in*], **inputs**(*varlist1*) **outputs**(*varlist2*) [*options*]

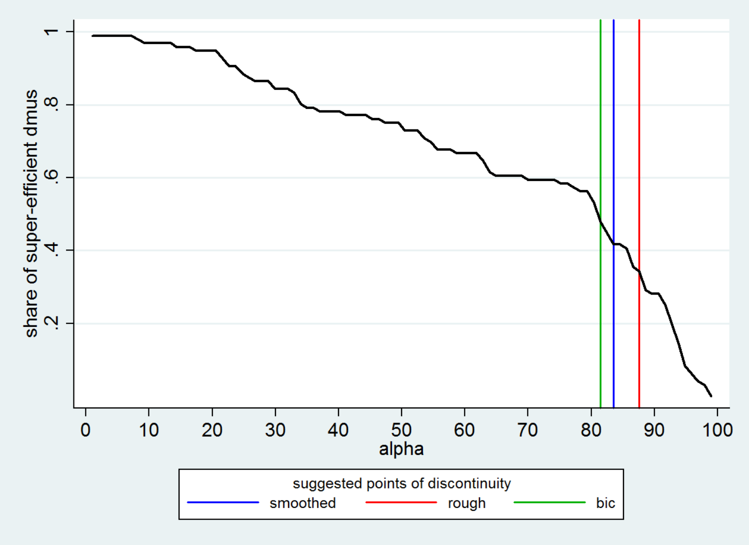
technology_definition	description
Model	
varname	identifier
inputs(*varlist1*)	list of input variables
outputs(*varlist2*)	list of output variables

options	description
Main	
ort(*string*)	*string* may be **output** or **input**; default is **ort**(*input*)
nalpha(*#*)	try *#* values for alpha; the maximum allowed value is *N* which is also the default
Detection	
nobic	do not suggest discontinuities based on BIC
norough	do not suggest discontinuities based on rough series of difference in differences
nosmooth	do not suggest discontinuities based on smoothed series of differences in differences
smoother(*string*)	use smoother *string* for smoothing series of differences in differences
Reporting	
noplot	suppress plotting series of share of super-efficient dmus
dots	display loop dots

weights are not allowed; see weight.
bootstrap, **by**, and **svy** are not allowed; see prefix.

Description

Graph. output oaoutlier: (regional health production in Bavaria)



oaoutlier: Saved Results

Viewer (#6) [help oaoutlier]

help oaoutlier

Advice | Contents | What's New | News

Examples

```
. oaoutlier firm, inputs(capital labor energy) outputs(durables perishables) ort(output) dots
```

Here *oaoutlier* and *orderalpha* precede *dea* and already respect *deas* requirements for naming variables:

```
. oaoutlier dmu, inputs(i_capital i_labor i_energy) outputs(o_durables o_perishables) ort(output)

. orderalpha dmu, inputs(i_capital i_labor i_energy) outputs(o_durables o_perishables) ort(output) invert alpha(`r(asmooth1)')
    gen(oaeffi)

. dea if oaeffi <= 1, rts(vrs) ort(o)
```

Saved results

orderalpha saves the following in **r()**:

Scalars

r(asmooth#)	point of discontinuity suggested by local rule using smoothed series (# may take values up to 3)
r(arough#)	point of discontinuity suggested by local rule using rough series (# may take values up to 3)
r(abic1)	point of discontinuity suggested by global rule
r(ssmooth#)	share of super-efficients dmus that corresponds to **r(***asmooth#***)**
r(srough#)	share of super-efficients dmus that corresponds to **r(***arough#***)**
r(sbic1)	share of super-efficients dmus that corresponds to **r(***abic1***)**

Macros

r(cmd)	**oaoutlier**
r(cmdline)	command as typed
r(title)	**Order-alpha based outlier detection**
r(ort)	either **input** or **output**

Matrices

r(oaresult)	*Nx2* matrix of values tried for alpha and corresponding shares of super-efficient dmus

References

Daraio, C. and L. Simar (2007). *Advanced robust and nonparametric methods in efficiency analysis: Methodology and applications.* Springer, New York.

Also see

Manual: **[R] smooth**

Help: **[R] smooth**

Online: **dea**, **orderalpha**

References

Aigner, D., Lovell, C.A.K. & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models, *Journal of Econometrics* **6**: 21–37.

Aragon, Y., Daouia, A. & Thomas-Agnan, C. (2005). Nonparametric frontier estimation: a conditional quantile based approach, *Econometric Theory* **21**: 358–389.

Cazals, C., Florens, J.P. & Simar, L. (2002). Nonparametric frontier estimation: A robust approach, *Journal of Econometrics* **106**: 1–25.

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

Daraio, C. & Simar, L. (2007). *Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications*, Springer, New York.

Deprins, D., Simar, L. & Tulkens, H. (1984). Measuring laborefficiency in post offices, in: Marchand, M., Pestieau, P. and Tulkens, H. (eds.), *The Performance of Public Enterprises: Concepts and Measurement*, Elsevier, Amsterdam, 243–267.

Ji, Y.B. & Lee, C. (2010). Data envelopment analysis, *The Stata Journal* **10**, 267–280.