The Stata Journal (2016)
16, Number 3, pp. 550–589

Nonparametric frontier analysis using Stata

Oleg Badunenko
University of Cologne
Cologne, Germany
obadunen@uni-koeln.de

Pavlo Mozharovskyi
Agrocampus Ouest
Rennes, France
pavlo.mozharovskyi@univ-rennes1.fr

Abstract. In this article, we describe five new Stata commands that fit and provide statistical inference in nonparametric frontier models. The `tenonradial` and `teradial` commands fit data envelopment models where nonradial and radial technical efficiency measures are computed (Färe, 1998, *Fundamentals of Production Theory*; Färe and Lovell, 1978, *Journal of Economic Theory* 19: 150–162; Färe, Grosskopf, and Lovell, 1994a, *Production Frontiers*). Technical efficiency measures are obtained by solving linear programming problems. The `teradialbc`, `nptestind`, and `nptestrts` commands provide tools for making statistical inference regarding radial technical efficiency measures (Simar and Wilson, 1998, *Management Science* 44: 49–61; 2000, *Journal of Applied Statistics* 27: 779–802; 2002, *European Journal of Operational Research* 139: 115–132). We provide a brief overview of the nonparametric efficiency measurement, and we describe the syntax and options of the new commands. Additionally, we provide an example showing the capabilities of the new commands. Finally, we perform a small empirical study of productivity growth.

Keywords: st0444, tenonradial, teradial, teradialbc, nptestind, nptestrts, nonparametric efficiency analysis, data envelopment analysis, technical efficiency, radial measure, nonradial measure, linear programming, bootstrap, subsampling bootstrap, smoothed bootstrap, bias correction, frontier analysis

1 Introduction

The concept of efficiency is at the core of production economics. Beginning with the pioneering work by Cobb and Douglas (1928), there have been many attempts to parameterize the production process, for example, via the Leontief constant elasticity of substitution, transcendental logarithmic production, and cost functions. Conceptually, researchers looked at the average input–output relationship assuming no inefficiency. However, it was not plausible to assume that all units are homogeneous, that is, operating at the same level of efficiency. Among the first to offer an appropriate modification was Farrell (1957), who built on the concept of efficiency postulated by Koopmans (1951) and Debreu (1951) and introduced a foundation that has become a distinct field in economics: the efficiency analysis. Färe (1998); Färe, Grosskopf, and Lovell (1994a); and Färe and Primont (1995) provide many insights into nonparametric efficiency measurement.

Data envelopment analysis (DEA), a leading analytical technique for measuring relative efficiency, has been widely used by both academic researchers and practitioners in evaluating the efficiency of decision-making units in terms of converting inputs into out-

 st0444

puts. Researchers choose this technique because it does not impose a priori functional form and it allows for multiple output technologies.

Although the DEA method is typically considered to be deterministic, the efficiency is still computed relative to the estimated frontier and not the true frontier. The efficiency scores obtained from a finite sample are subject to sampling variation of the estimated frontier. Simar and Wilson (1998, 2000, 2002) have laid out a statistical model and proposed consistent bootstrap procedures to provide statistical inference regarding technical efficiency measures in nonparametric frontier models.

The estimation of DEA models can be readily performed in Stata with the user-written command `dea` (Ji and Lee 2010). However, `dea` is limited in its capability and is slow with even moderate datasets. We provide a time comparison of `dea` and our commands. The five new Stata commands described here fit and provide statistical inference in nonparametric frontier models. `tenonradial` and `teradial` fit data envelopment models where nonradial and radial technical efficiency measures are computed (Färe 1998; Färe and Lovell 1978; Färe, Grosskopf, and Lovell 1994a). `teradialbc`, `nptestind`, and `nptestrts` provide tools to make statistical inference regarding radial technical efficiency measures (Simar and Wilson 1998, 2000, 2002).

The remainder of this article is structured as follows: section 2 provides an overview of nonparametric frontier models; sections 3–7 contain the syntax and explain the options of the new commands; section 8 illustrates the capabilities of the new commands using a dataset for program follow-through at 70 U.S. primary schools and performs the analysis of the changes in productivity for 52 countries using Penn World Tables; section 9 details the features and limitations of the new commands; section 10 emphasizes the differences between our commands and the `dea` command; and section 11 concludes the article.

2 Nonparametric frontier analysis

In this section, we introduce two types of nonparametric efficiency measurement: radial and nonradial. We also discuss recent statistical developments regarding radial measures. The exposition here is only introductory. For more details, refer to the cited works.

2.1 Radial efficiency analysis

Our measures of technical efficiency for the production data points are conventional radial Debreu–Farrell measures of efficiency loss (Debreu 1951; Farrell 1957). For each data point k ($k = 1, \dots, K$), vector $x_k = (x_{k1}, \dots, x_{kN}) \in \Re^N$ denotes N inputs and vector $y_k = (y_{k1}, \dots, y_{kM}) \in \Re^M$ denotes M outputs. We assume that under technology T the data (y, x) are such that outputs are producible by inputs,

$$T = \{(x, y) : y \text{ are producible by } x\} \tag{1}$$

The technology is fully characterized by its production possibility set,

$$P(x) \equiv \{y : (x,y) \in T\} \tag{2}$$

or input requirement set,

$$L(y) \equiv \{x : (x,y) \in T\} \tag{3}$$

Conditions (2) and (3) imply that the available outputs and inputs are feasible. The upper boundary of the production possibility set and lower boundary of the input requirement set define the frontier. How far a given data point is from the frontier represents its efficiency. In output-based radial efficiency measurement, the amount of necessary (proportional) expansion of outputs to move a data point to a boundary of the production possibility set $P(x)$ serves as a measure of technical efficiency. In input-based radial efficiency measurement, it is instead the amount of necessary (proportional) reduction of inputs to move a data point to a boundary of the input requirement set $L(y)$.

Empirically, technical efficiencies are fit via activity analysis models, widely known as DEA models. For K data points, M outputs, and N inputs, an estimate of the radial Debreu–Farrell output-based measure of technical efficiency can be calculated by solving a linear programming problem for each data point k ($k = 1, \dots, K$):

$$\begin{aligned} \widehat{F}_k^o(y_k, x_k, y, x|\text{CRS}) = \max_{\theta, z} \theta \\ \text{s.t. } \sum_{k=1}^{K} z_k y_{km} \geq y_{km}\theta_m, m = 1, \dots, M \\ \sum_{k=1}^{K} z_k x_{kn} \leq x_{kn}, n = 1, \dots, N \\ z_k \geq 0 \end{aligned} \tag{4}$$

y is a $K \times M$ matrix of available data on outputs, and x is a $K \times N$ matrix of available data on inputs. The estimate of $P(x)$ is the smallest convex free-disposal hull that envelops the observed data, the upper boundary of which is a piecewise linear estimate of the true best-practice frontier of $P(x)$. Equation (4) gives us constant returns to scale (CRS) specification. Other returns to scale are modeled by adjusting process operating levels z_k: for variable returns to scale (VRS), a convexity constraint $\sum_{k=1}^{K} z_k = 1$ is added,[1] while for nonincreasing returns to scale (NIRS), the $\sum_{k=1}^{K} z_k \leq 1$ inequality is added[2] to the set of restrictions in the linear programming problem in (4).

1. This equality ensures that data point k is compared only with data points of similar size; under the CRS assumption, data points of different sizes might be compared with one another.
2. This inequality ensures that data point k is not compared with other data points that are considerably larger. It may be compared with smaller data points.

To facilitate the discussion, figures 1 and 2 present hypothetical one-input one-output production processes with three different technologies: CRS, VRS, and NIRS. Conceptually, in figure 1 (figure 2) the vertical (horizontal) distance from a data point (x_i, y_i) or (x_j, y_j) to the CRS, VRS, and NIRS best-practice frontier stands for output-based (input-based) technical efficiency under the assumption of CRS, VRS, and NIRS technology. In a multidimensional case, the required distance is the radial path from a data point that is parallel to axes along which all outputs (inputs) are measured.

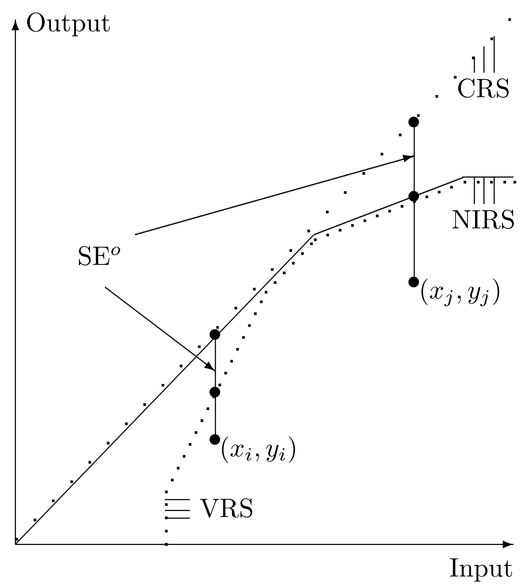

Figure 1. Output-based technical and scale efficiency

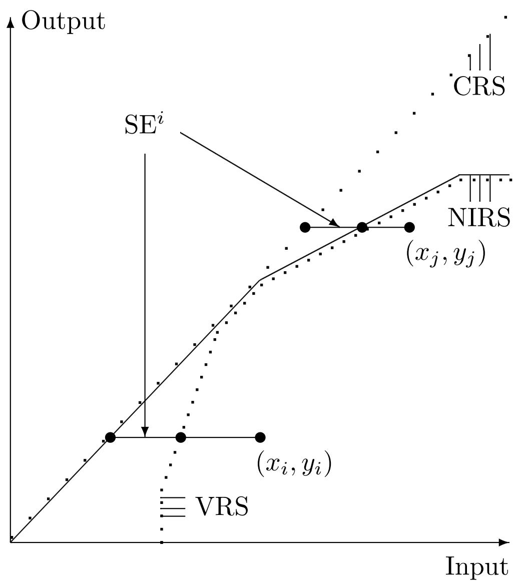

Figure 2. Input-based technical and scale efficiency

2.2 Nonradial efficiency analysis

For data point (y_k, x_k), the radial measure expands (shrinks) all M outputs ($y_k = y_{k1}, \ldots, y_{kM}$) and N inputs ($x_k = x_{k1}, \ldots, x_{kN}$) proportionally until the frontier is reached. At the reached frontier point, some but not all outputs (inputs) can be expanded (shrunk) while remaining feasible. If such possibility is available for a given data point k for output m (input n), then the reference point $F_k^o(y_k, x_k) \times y_{mk}$ $[F_k^i(y_k, x_k) \times x_{nk}]$ is said to have slack in output y_m (input x_n). A nonradial measure of technical efficiency, the Russell measure (RM), accommodates such slacks (Färe and Lovell 1978; Färe, Grosskopf, and Lovell 1994a). The output-based nonradial measure for data point j is defined by

$$\mathrm{RM}_k^o(y_k, x_k, y, x|\mathrm{CRS}) = \max \left\{ \begin{array}{r} M^{-1} \sum\limits_{m=1}^{M} \theta_m\text{:} \;\; (\theta_1 y_{k1}, \ldots, \theta_M y_{kM}) \in P(x), \\ \theta_m \geq 0, m = 1, \ldots, M \end{array} \right\}$$

The input-based counterpart is given by

$$\mathrm{RM}_k^i(y_k, x_k, y, x|\mathrm{CRS}) = \min \left\{ \begin{array}{r} N^{-1} \sum\limits_{n=1}^{N} \lambda_n\text{:} \;\; (\lambda_1 x_{k1}, \ldots, \lambda_N y_{kN}) \in L(y), \\ \lambda_n \geq 0, n = 1, \ldots, N \end{array} \right\}$$

The output-based RM can be calculated for positive outputs as a solution to the linear programming problem

$$\widehat{\text{RM}}_k^o(y_k, x_k, y, x|\text{CRS}) = M^{-1} \max_{\theta,z} \sum_{m=1}^{M} \theta_m \tag{5}$$
$$\text{s.t.} \quad \sum_{k=1}^{K} z_k y_{km} \geq y_{km}\theta_m, m = 1, \ldots, M$$
$$\sum_{k=1}^{K} z_k x_{kn} \leq x_{kn}, n = 1, \ldots, N$$
$$z_k \geq 0$$

and the input-based RM can be calculated for positive inputs as a solution to the linear programming problem

$$\widehat{\text{RM}}_k^i(y_k, x_k, y, x|\text{CRS}) = N^{-1} \min_{\theta,z} \sum_{n=1}^{N} \lambda_n \tag{6}$$
$$\text{s.t.} \quad \sum_{k=1}^{K} z_k y_{km} \geq y_{km}, m = 1, \ldots, M$$
$$\sum_{k=1}^{K} z_k x_{kn} \leq x_{kn}\lambda_n, n = 1, \ldots, N$$
$$z_k \geq 0$$

If output $y_{km} = 0$ ($x_{kn} = 0$), then the linear programming problem in (5) [in (6)] is modified and θ_m (λ_n) is set to 1.

The RM allows for nonproportional expansions (reductions) in each positive output (input). The nonradial output-based (input-based) RM collapses to the radial measure when $\theta_m = \theta, \forall m$, where $y_{km} > 0$ ($\lambda_n = \lambda, \forall n$, where $x_{kn} > 0$). However, because the RM can expand (shrink) an output (input) vector at most (least) as far as the radial measure can, we have the result that

$$1 \geq \widehat{F}_k^o(y_k, x_k, y, x|\text{CRS}) \geq \widehat{\text{RM}}_k^o(y_k, x_k, y, x|\text{CRS})$$

and

$$0 < \widehat{\text{RM}}_k^i(y_k, x_k, y, x|\text{CRS}) \leq \widehat{F}_k^i(y_k, x_k, y, x|\text{CRS}) \leq 1$$

Technologies under NIRS and VRS can be modeled by imposing respective restrictions on the intensity vector, z, in the piecewise linear technology, that is, in (5) and (6). Then the RM can be calculated relative to these technologies.

With just one input (output), the input-based (output-based) RM is equal to the Debreu–Farrell radial measure of technical efficiency.

2.3 Statistical inference in the radial frontier model

Although the DEA method is typically considered to be deterministic, the efficiency is still computed relative to the estimated frontier and not the true frontier. The efficiency scores obtained from a finite sample [in (4) from K data points] are subject to sampling variation of the estimated frontier. The estimated technical efficiency measures are too optimistic, caused by the DEA estimate of the production set necessarily being a weak subset of the true production set under standard assumptions underlying DEA. The statistical inference regarding the radial DEA estimates can be provided via the bootstrap technique. The details of the concept and implementation of the bootstrap mechanism are given in Simar and Wilson (1998, 2000) and Kneip, Simar, and Wilson (2008). The bootstrapping procedure allows estimation of the bias and the confidence interval of the original estimate. Badunenko, Henderson, and Kumbhakar (2012) study statistical properties of the bias-corrected estimator in finite samples.

2.4 Type of bootstrap for statistical inference

The bootstrapping technique mentioned in the previous section relies on several assumptions. In output-based efficiency measurement, the major assumption depends on whether the estimated output-based measures of technical efficiency are independent of the mix of outputs. In input-based efficiency measurement, the major assumption depends on whether the estimated input-based measures of technical efficiency are independent of the mix of inputs. This dependency is testable given the assumption of returns to scale of the global technology (Wilson 2003). If output-based measures of technical efficiency are independent of the mix of outputs, the smoothed homogeneous bootstrap can be used. This type of bootstrap is not computer intensive. If, on the contrary, output-based measures of technical efficiency are not independent of the mix of outputs, then the heterogeneous bootstrap must be used to provide valid statistical inference. The latter type of bootstrap is quite demanding computationally and may take a while for large datasets.

2.5 Returns to scale and scale analysis

The assumption regarding the global technology is crucial in DEA. Depending on this assumption, (4) and resulting measures of technical efficiency will vary. The assumption about returns to scale should be made using prior knowledge about the particular industry. If this knowledge does not suffice or is not conclusive, then the returns to scale assumption can be tested econometrically. Moreover, if technology is not CRS globally, then estimating the measure of technical efficiency under CRS will lead to inconsistent results (Simar and Wilson 2002).

The measures of radial technical efficiency in (4) under CRS, NIRS, and VRS can be used to calculate the measures of scale efficiency, originally proposed by Färe and Grosskopf (1985),

$$S_k^o(y_k, x_k) = \frac{\widehat{F}_k^o(y_k, x_k, y, x|\text{CRS})}{\widehat{F}_k^o(y_k, x_k, y, x|\text{VRS})} \tag{7}$$

and

$$S_k^{o*}(y_k, x_k) = \frac{\widehat{F}_k^o(y_k, x_k, y, x|\text{NIRS})}{\widehat{F}_k^o(y_k, x_k, y, x|\text{VRS})}$$

for output-based analysis, and

$$S_k^i(y_k, x_k) = \frac{\widehat{F}_k^i(y_k, x_k, y, x|\text{CRS})}{\widehat{F}_k^i(y_k, x_k, y, x|\text{VRS})}$$

and

$$S_k^{i*}(y_k, x_k) = \frac{\widehat{F}_k^i(y_k, x_k, y, x|\text{NIRS})}{\widehat{F}_k^i(y_k, x_k, y, x|\text{VRS})}$$

for input-based analysis. Scale efficiency S_k^o measures how close the data point (y_k, x_k) is to potentially optimal scale, also known as maximum productive scale size, which is the portion of the frontier where the CRS and VRS frontiers coincide in figures 1 and 2 (denoted by SE^o and SE^i, respectively). If $S_k^o(y_k, x_k) = 1$ [$S_k^i(y_k, x_k) = 1$ in input-based efficiency measurement], then a data point (y_k, x_k) is scale efficient. If $S_k^o(y_k, x_k) > 1$ [$S_k^i(y_k, x_k) < 1$ in input-based efficiency measurement], then a data point (y_k, x_k) is scale inefficient because it operates under the decreasing returns portion of technology if $S_k^{o*}(y_k, x_k) = 1$ [$S_k^{i*}(y_k, x_k) = 1$ in input-based efficiency measurement] or because it operates under the increasing returns portion of technology if $S_k^{o*}(y_k, x_k) > 1$ [$S_k^{i*}(y_k, x_k) < 1$ in input-based efficiency measurement].

On one hand, if global technology T in (1) represents CRS, then the VRS estimator is less efficient than CRS. On the other hand, if global technology T in (1) is not CRS at some mix of outputs (inputs), then the CRS estimator is inconsistent. Therefore, Simar and Wilson (2002) suggest the following tests:

$$\begin{aligned} \text{Test \#1:} \quad & H_0 : T \text{ is globally CRS} \\ & H_1 : T \text{ is VRS} \end{aligned}$$

If null hypothesis H_0 is rejected (that is, if technology is not CRS everywhere), then the following test with a less restrictive null hypothesis may be performed:

$$\begin{aligned} \text{Test \#2:} \quad & H_0' : T \text{ is globally NIRS} \\ & H_1 : T \text{ is VRS} \end{aligned}$$

Using scale efficiency measures for all K data points, the statistics for testing test #1 and test #2 are defined by

$$\widehat{S}_{2n}^o = \frac{\sum_{k=1}^K \widehat{F}_k^o(y_k, x_k, y, x|\text{CRS})}{\sum_{k=1}^K \widehat{F}_k^o(y_k, x_k, y, x|\text{VRS})} \tag{8}$$

and

$$\widehat{S}_{2n}^{o\prime} = \frac{\sum_{k=1}^{K} \widehat{F}_k^o(y_k, x_k, y, x|\text{NIRS})}{\sum_{k=1}^{K} \widehat{F}_k^o(y_k, x_k, y, x|\text{VRS})} \tag{9}$$

The idea of testing the null hypothesis that the technology is globally CRS versus the alternative hypothesis that the technology is globally VRS, test #1, boils down to testing how far the test statistic (8) is from its bootstrap analog. This statistic represents the ratio of the average measures of technical efficiency under the assumption of VRS and CRS technologies. If the null hypothesis is true, then the average distance between the VRS and CRS frontiers is small. If the alternative hypothesis is true, then the average distance between the VRS and CRS frontiers is large—the null hypothesis H_0 is rejected if $\widehat{S}_{2n}^o$ is significantly larger than 1 [if $\widehat{S}_{2n}^i$, defined similarly to (8), is smaller than 1 in input-based efficiency measurement]. If H_0 is rejected, then test #2 can be performed to test the null hypothesis H_0' that the technology is globally NIRS versus the alternative hypothesis that the technology is globally VRS. Analogously to test #1, if the null hypothesis H_0' is true, then the average distance between the VRS and NIRS frontiers is small. If the alternative hypothesis is true, then the average distance between the VRS and NIRS frontiers is large—the null hypothesis H_0' is rejected if $\widehat{S}_{2n}^{o\prime}$ is significantly larger than 1 [if $\widehat{S}_{2n}^{i\prime}$, defined similarly to (9), is smaller than 1 in input-based efficiency measurement].

Because of the importance of the returns to scale assumption for the DEA estimator, this data-driven test should be performed before applying any DEA model.

Additionally, this testing procedure can be used to perform the scale analysis for each data point. The CRS assumption is only feasible when all data points are operating at an optimal scale, that is, when scale efficiency is unity. However, for many reasons (for example, imperfect competition or financial constraints), it is more appropriate to assume VRS (see Coelli, Rao, O'Donnell, and Battese [2005] for history and development of this stream). Assuming CRS when VRS should be assumed in reality overestimates the technical efficiency estimate exactly by scale efficiency. Therefore, performing the individual returns to scale test is fairly important in the case of scale efficiency analysis.

First, for each data point k, the null hypothesis of test $\#1_k$ that measures of technical efficiency are equal under CRS and VRS, or $S_k^o(y_k, x_k) = 1$, against the alternative hypothesis that $S_k^o(y_k, x_k) > 1$ [$S_k^i(y_k, x_k) < 1$ in input-based case] is tested.[3] Because by definition $S_k^o(y_k, x_k) \geq 1$ [$S_k^i(y_k, x_k) \geq 1$ in input-based case], this null hypothesis is rejected if $S_k^o(y_k, x_k)$ is significantly greater than 1 [$S_k^i(y_k, x_k) \leq 1$ in input-based case]. The data point $S_k^o(y_k, x_k)$, for which this null hypothesis is rejected, $S_k^o(y_k, x_k) > 1$ [$S_k^i(y_k, x_k) < 1$ in input-based case], is said to be scale inefficient.

Second, for all scale inefficient data points, the null hypothesis of test $\#2_k$ that the measures of technical efficiency are equal under NIRS and VRS, or $S_k^{o*}(y_k, x_k) = 1$ [$S_k^{i*}(y_k, x_k) = 1$ in input-based case], against the alternative that $S_k^{o*}(y_k, x_k) > 1$

3. Note that $S_k^o(y_k, x_k)$ is the test statistic of test $\#1_k$.

[$S_k^{i*}(y_k, x_k) < 1$ in input-based case] can be performed. Test $\#2_k$ concludes that data point (y_k, x_k) is operating under increasing returns to scale [such as a data point (x_i, y_i) in terms of figure 1 or 2] if $S_k^{o*}(y_k, x_k)$ is significantly larger than 1 [$S_k^{i*}(y_k, x_k) < 1$ in input-based case], or is operating under decreasing returns to scale [such as a data point (x_j, y_j) in terms of figure 1] otherwise. All tests in this subsection are based on bootstrap techniques mentioned in the previous section.

3 The tenonradial command

`tenonradial` uses reduced linear programming to compute the nonradial output- or input-based measure of technical efficiency, which is known as the RM. In input-based nonradial efficiency measurement, the RM allows for nonproportional (different) reductions in each positive input, and this is what permits it to shrink an input vector all the way back to the efficient subset. In output-based nonradial efficiency measurement, the RM allows for nonproportional (different) expansions of each positive output.

3.1 Syntax

`tenonradial` *outputs* `=` *inputs* [`(`*ref_outputs* `=` *ref_inputs*`)`] [*if*] [*in*] [`,`
`rts(`*rtsassumption*`)` `base(`*basetype*`)` `reference(`*varname*`)` `tename(`*newvar*`)`
`noprint`]

Specification

outputs is the list of output variables.

inputs is the list of input variables.

ref_outputs is the optional list of output variables for the reference set. The number of variables in *ref_outputs* must be equal to the number of variables in *outputs*. If *ref_outputs* is specified, then *ref_inputs* must also be specified.

ref_inputs is the optional list of input variables for the reference set. The number of variables in *ref_inputs* must be equal to the number of variables in *inputs*. If *ref_inputs* is specified, then *ref_outputs* must also be specified.

3.2 Options

Technology

`rts(`*rtsassumption*`)` specifies the returns to scale assumption.

Specifying `rts(crs)` requests that the measure of technical efficiency be computed under the assumption of CRS. This is the default.

Specifying `rts(nirs)` requests that the measure of technical efficiency be computed under the assumption of NIRS.

Specifying `rts(vrs)` requests that the measure of technical efficiency be computed under the assumption of VRS.

`base(`*basetype*`)` specifies the type of optimization.

Specifying `base(output)` requests that the output-based measure be computed. This is the default.

Specifying `base(input)` requests that the input-based measure be computed.

Reference set

`reference(`*varname*`)` specifies the indicator variable that defines which data points of *outputs* and *inputs* form the technology reference set. If *ref_outputs* and *ref_inputs* are specified, then *varname* defines which data points of *ref_outputs* and *ref_inputs* form the technology reference set.

Variable generation

`tename(`*newvar*`)` creates *newvar* containing the nonradial measures of technical efficiency.

Miscellaneous

`noprint` suppresses the estimation details, description of the data, and reference set.

3.3 Output, generated variable, and stored results

If `noprint` is not specified, then `tenonradial` produces the summary of the model, the data, and a note about the reference set. Specifying `tename(`*newvar*`)` will generate *newvar* containing the nonradial measures of technical efficiency in the current dataset.

`tenonradial` stores the following in `e()`:

Scalars	
`e(M)`	number of outputs
`e(N)`	number of inputs
`e(K)`	number of data points
`e(Kref)`	number of data points for the reference set
Macros	
`e(cmd)`	`tenonradial`
`e(cmdline)`	command as typed
`e(title)`	title in estimation output
`e(rts)`	`CRS`, `NIRS`, or `VRS`
`e(base)`	`output` or `input`
`e(outputs)`	list of output variables
`e(inputs)`	list of input variables
`e(ref_outputs)`	list of output variables for the reference set
`e(ref_inputs)`	list of input variables for the reference set
Matrices	
`e(te)`	$K \times 1$ matrix with measures of technical efficiency
Functions	
`e(sample)`	marks estimation sample

4 The teradial command

The syntax, options, output, generated variable, and stored results are identical to those of `tenonradial`.

5 The teradialbc command

`teradialbc` performs statistical inference about the radial measure of technical efficiency.

5.1 Syntax

`teradialbc` *outputs* `=` *inputs* [`(`*ref_outputs* `=` *ref_inputs*`)`] [*if*] [*in*] [`,`
`rts(`*rtsassumption*`)` `base(`*basetype*`)` `reference(`*varname*`)` `subsampling`
`kappa(`*#*`)` `smoothed` `heterogeneous` `reps(`*#*`)` `level(`*#*`)` `tename(`*newvar*`)`
`tebc(`*newvar*`)` `biasboot(`*newvar*`)` `varboot(`*newvar*`)` `biassqvar(`*newvar*`)`
`telower(`*newvar*`)` `teupper(`*newvar*`)` `noprint` `nodots`]

Specification

The “specifications” are identical to those of `tenonradial`.

5.2 Options

Technology

The “technology” options are identical to those of `tenonradial`.

Reference set

The “reference set” option is identical to that of `tenonradial`.

Bootstrap

`subsampling` requests that the reference set be bootstrapped with subsampling. If `subsampling` is not specified, then bootstrap with smoothing is used.

`kappa(`#`)` sets the size of the subsample as $K^{\texttt{kappa}}$, where K is the number of data points in the original reference set. The default is `kappa(0.7)`. # may be between 0.5 and 1.

`smoothed` requests that the reference set be bootstrapped with smoothing. This is the default. This option is for keeping track of the bootstrap type.

`heterogeneous` requests that the reference set be bootstrapped with heterogeneous smoothing. If `heterogeneous` is not specified, then bootstrap with homogeneous smoothing is used.

`reps(`#`)` specifies the number of bootstrap replications to be performed. The default is `reps(999)`; the minimum is `reps(200)`. Adequate estimates of confidence intervals using bias-corrected methods typically require 1,000 or more replications.

Statistical inference

`level(`#`)` sets the confidence level; the default is `level(95)`.

Variable generation

`tename(`*newvar*`)` creates *newvar* containing the radial measures of technical efficiency.

`tebc(`*newvar*`)` creates *newvar* with the bias-corrected radial measures of technical efficiency.

`biasboot(`*newvar*`)` creates *newvar* with the bootstrap bias estimate for the original radial measures of technical efficiency.

`varboot(`*newvar*`)` creates *newvar* with the bootstrap variance estimate for the radial measures of technical efficiency.

`biassqvar(`*newvar*`)` creates *newvar* with three times the ratio of bias squared to variance for the radial measures of technical efficiency.

`telower(`*newvar*`)` creates *newvar* with the lower-bound estimate for the radial measures of technical efficiency.

`teupper(`*newvar*`)` creates *newvar* with the upper-bound estimate for the radial measures of technical efficiency.

Miscellaneous

`noprint` suppresses the estimation details, description of the data, and reference set.

`nodots` suppresses display of the replication dots. By default, one dot character is displayed for each successful replication. A red "x" is displayed if the command returns an error.

5.3 Details

`teradialbc` performs bias correction of the radial output- or input-based measure of technical efficiency under the assumption of CRS, NIRS, or VRS technology. It also computes bias and constructs confidence intervals.

If a reference set is not specified, then the reference set is formed by data points for which measures of technical efficiency are computed.

Statistical inference (computation of bias, variance, and confidence interval) is performed for data points where the real number of bootstrap replications is at least 100. Matrix `e(realreps)` stores the real number of bootstrap replications, which may be smaller than `reps(`#`)`.

If at least one input-based bias-corrected Farrell measure of technical efficiency is negative, then the analysis and statistical inference is performed in terms of Shephard distance functions, a reciprocal of the Debreu–Farrell measure.

5.4 Dependency of teradialbc

`teradialbc` depends on the Mata functions `kdens_bw()` and `mm_quantile()`. If not already installed, you can install these functions by typing `ssc install kdens` and `ssc install moremata`, respectively.

5.5 Output, generated variables, and stored results

If `noprint` and `nodots` are not specified, then `teradialbc` produces the summary of the model, the data, and a note about the reference set, and it displays replication dots. Several variables related to statistical inference can be generated in the current dataset by specifying options. For example, specifying `tename(`*newvar*`)` will generate *newvar* containing the radial measures of technical efficiency.

teradialbc stores the following in e():

Scalars	
e(M)	number of outputs
e(N)	number of inputs
e(K)	number of data points
e(Kref)	number of data points for the reference set
e(reps)	number of bootstrap replications
Macros	
e(cmd)	teradialbc
e(cmdline)	command as typed
e(title)	title in estimation output
e(rts)	CRS, NIRS, or VRS
e(base)	output or input
e(outputs)	list of output variables
e(inputs)	list of input variables
e(ref_outputs)	list of output variables for the reference set
e(ref_inputs)	list of input variables for the reference set
Matrices	
e(te)	$K \times 1$ matrix with measures of technical efficiency
e(tebc)	$K \times 1$ matrix with bias-corrected radial measures of technical efficiency
e(biasboot)	$K \times 1$ matrix with bootstrap bias estimate for original radial measures of technical efficiency
e(varboot)	$K \times 1$ matrix with bootstrap variance estimate for radial measures of technical efficiency
e(telow)	$K \times 1$ matrix with lower-bound estimate for radial measures of technical efficiency
e(teupp)	$K \times 1$ matrix with upper-bound estimate for radial measures of technical efficiency
e(biassqvar)	$K \times 1$ matrix with three times the ratio of bias squared to variance for radial measures of technical efficiency
e(realreps)	$K \times 1$ matrix with number of bootstrap replications used for statistical inference
Functions	
e(sample)	marks estimation sample

6 The nptestind command

nptestind performs nonparametric tests of independence.

6.1 Syntax

nptestind *outputs* = *inputs* [*if*] [*in*] [, rts(*rtsassumption*) base(*basetype*) reps(#) alpha(#) noprint nodots]

Specifications

outputs is the list of output variables.

inputs is the list of input variables.

6.2 Options

Technology

The "technology" options are identical to those of `tenonradial`.

Bootstrap

`reps(#)` specifies the number of bootstrap replications to be performed. The default is `reps(999)`; the minimum is `reps(200)`. Adequate estimates of confidence intervals using bias-corrected methods typically require 1,000 or more replications.

Statistical inference

`alpha(#)` sets the significance level. The default is `alpha(0.05)`.

Miscellaneous

The "miscellaneous" options are identical to those of `teradialbc`.

6.3 Dependency of nptestind

`nptestind` depends on the Mata function `kdens_bw()`. If not already installed, you can install this function by typing `ssc install kdens`.

6.4 Output and stored results

If `noprint` and `nodots` are not specified, then `nptestind` produces the summary of the model, the data, and a note about the reference set, and it displays replication dots.

`nptestind` stores the following in `e()`:

Scalars	
`e(M)`	number of outputs
`e(N)`	number of inputs
`e(K)`	number of data points
`e(pvalue)`	p-value of the test that the measure of technical efficiency and the mix of inputs (or outputs) are independent
`e(t4n)`	T4n statistic
`e(reps)`	number of bootstrap replications
Macros	
`e(cmd)`	`nptestind`
`e(cmdline)`	command as typed
`e(title)`	title in estimation output
`e(rts)`	`CRS`, `NIRS`, or `VRS`
`e(base)`	`output` or `input`
`e(outputs)`	list of output variables
`e(inputs)`	list of input variables
Matrices	
`e(t4nboot)`	reps $\times$ 1 matrix with bootstrap values of the T4n statistic

7 The nptestrts command

`nptestrts` performs nonparametric tests of returns to scale.

7.1 Syntax

`nptestrts` *outputs* `=` *inputs* [*if*] [*in*] [`, base(`*basetype*`) heterogeneous reps(`*#*`) alpha(`*#*`) testtwo tecrsname(`*newvar*`) tenirsname(`*newvar*`) tevrsname(`*newvar*`) sefficiency(`*newvar*`) psefficient(`*newvar*`) sefficient(`*newvar*`) nrsovervrs(`*newvar*`) pineffdrs(`*newvar*`) sineffdrs(`*newvar*`) noprint nodots`]

Specifications

The "specifications" are identical to those of `nptestind`.

7.2 Options

Technology

`base(`*basetype*`)` specifies the type of optimization.

Specifying `base(output)` requests that the output-based measure be computed. This is the default.

Specifying `base(input)` requests that the input-based measure be computed.

Bootstrap

`heterogeneous` requests that the reference set be bootstrapped with heterogeneous smoothing. If `heterogeneous` is not specified, then bootstrap with homogeneous smoothing is used.

`reps(`*#*`)` specifies the number of bootstrap replications to be performed. The default is `reps(999)`; the minimum is `reps(200)`. Adequate estimates of confidence intervals using bias-corrected methods typically require 1,000 or more replications.

Statistical inference

`alpha(`*#*`)` sets the significance level. The default is `alpha(0.05)`.

`testtwo` specifies that test #2 be performed.

If `testtwo` is not specified, then `nptestrts` performs only test #1, which consists of two parts. First, the null hypothesis that the technology is globally CRS (versus VRS) is tested. Second, the null hypothesis that the data point is scale efficient is tested.

If `testtwo` is specified, then `nptestrts` may also perform test #2. If the null hypothesis that the technology is CRS is rejected, then `testtwo` requests that `nptestrts` tests the null hypothesis that the technology is NIRS (versus VRS). If not all data points are scale efficient, then `nptestrts` tests that the reason for scale inefficiency is operating under decreasing returns to scale (DRS). If the null hypothesis that the technology is CRS is not rejected and all data points are scale efficient, then `nptestrts` will not perform test #2 even if `testtwo` is specified.

Variable generation

`tecrsname(`*newvar*`)` creates *newvar* containing the radial measures of technical efficiency under the assumption of CRS.

`tenirsname(`*newvar*`)` creates *newvar* containing the radial measures of technical efficiency under the assumption of NIRS.

`tevrsname(`*newvar*`)` creates *newvar* containing the radial measures of technical efficiency under the assumption of VRS.

`sefficiency(`*newvar*`)` creates *newvar* containing scale efficiency, the ratio of the measures of technical efficiency under CRS and VRS.

`psefficient(`*newvar*`)` creates *newvar* containing the *p*-value of the test that the data point is statistically scale efficient.

`sefficient(`*newvar*`)` creates indicator *newvar* equal to 1 if the data point is statistically scale efficient.

`nrsovervrs(`*newvar*`)` creates *newvar* containing the ratio of the measures of technical efficiency under NIRS and VRS.

`pineffdrs(`*newvar*`)` creates *newvar* containing the *p*-value of the test that the data point is scale inefficient due to operating under DRS.

`sineffdrs(`*newvar*`)` creates indicator *newvar* equal to 1 if the data point is statistically scale inefficient due to operating under DRS.

Miscellaneous

`noprint` suppresses the estimation details, description of the data, and reference set.

`nodots` suppresses display of the replication dots. One dot character is displayed for each successful replication.

7.3 Details

`nptestrts` performs nonparametric tests of returns to scale.

If `testtwo` is not specified, then `nptestrts` performs only test #1, which consists of two parts. First, the null hypothesis that the technology is globally CRS (versus VRS) is tested. Second, the null hypothesis that the data point is scale efficient is tested.

If `testtwo` is specified, then `nptestrts` may also perform test #2. If the null hypothesis that the technology is CRS is rejected, then `testtwo` requests that `nptestrts` tests the null hypothesis that the technology is NIRS (versus VRS). If not all data points are scale efficient, then `nptestrts` tests that the reason for scale inefficiency is operating under DRS. If the null hypothesis that the technology is CRS is not rejected and all data points are scale efficient, then `nptestrts` will not perform test #2 even if `testtwo` is specified.

7.4 Dependency of nptestrts

`nptestrts` depends on the Mata function `kdens_bw()`. If not already installed, you can install this function by typing `ssc install kdens`.

7.5 Output, generated variables, and stored results

If `noprint` and `nodots` are not specified, then `nptestrts` produces the summary of the model, the data, and a note about the reference set, and it displays replication dots. Several variables related to nonparametric tests can be generated in the current dataset by specifying options. For example, specifying `tecrsname(`*newvar*`)` will generate *newvar* containing the radial measures of technical efficiency under the assumption of CRS.

`nptestrts` stores the following in `e()`:

Scalars	
`e(M)`	number of outputs
`e(N)`	number of inputs
`e(K)`	number of data points
`e(sefficiencyMean)`	ratio of means of technical efficiency measures under CRS and VRS
`e(nsefficient)`	number of scale efficient data points
`e(nrsOVERvrsMean)`	ratio of means of technical efficiency measures under NIRS and VRS (if `testtwo`)
`e(pGlobalCRS)`	*p*-value of test that technology is globally CRS
`e(pGlobalNRS)`	*p*-value of test that technology is globally NIRS (if `testtwo`)
`e(reps)`	number of bootstrap replications
Macros	
`e(cmd)`	`nptestrts`
`e(cmdline)`	command as typed
`e(title)`	title in estimation output
`e(base)`	`output` or `input`
`e(outputs)`	list of output variables
`e(inputs)`	list of input variables
`e(smoothtype)`	`homogeneous` or `heterogeneous`

Matrices

`e(tecrsname)`	$K \times 1$ matrix with measures of technical efficiency under the assumption of CRS
`e(tenirsname)`	$K \times 1$ matrix with measures of technical efficiency under the assumption of NIRS
`e(tevrsname)`	$K \times 1$ matrix with measures of technical efficiency under the assumption of VRS
`e(sefficiency)`	$K \times 1$ matrix containing scale efficiency
`e(psefficient)`	$K \times 1$ matrix containing p-value of the test that the data point is statistically scale efficient
`e(sefficient)`	$K \times 1$ matrix containing 1s if statistically scale efficient
`e(nrsovervrs)`	$K \times 1$ matrix containing ratio of measures of technical efficiency under NIRS and VRS (if `testtwo`)
`e(pineffdrs)`	$K \times 1$ matrix containing p-value of the test that the data point is scale inefficient due to DRS (if `testtwo`)
`e(sineffdrs)`	$K \times 1$ matrix containing 1s if statistically scale inefficient due to DRS (if `testtwo`)

Functions

`e(sample)`	marks estimation sample

8 Empirical application

In this section, we show how to use the new commands and interpret the output based on two widely used datasets.

8.1 Data: CCR81

The first dataset comes from Charnes, Cooper, and Rhodes (1981). The data were originally used to evaluate the efficiency of public programs and their management. In what follows, we stick to output-based efficiency measurement.

We artificially create a variable, `dref`, to illustrate the capabilities of the new commands. We do not suppress the estimation details, description of the data, and reference set for the output-based radial measure of technical efficiency under the assumption of CRS technology. (We do suppress these for the remaining radial and all the nonradial measures.) Finally, we list the measures for the first seven observations:

```
. set seed 717117
. use ccr81
(Program Follow Through at 70 US Primary Schools)
. generate dref = x5 != 10
. teradial y1 y2 y3 = x1 x2 x3 x4 x5, rts(crs) base(output) reference(dref)
> tename(TErdCRSo)
 Radial (Debreu-Farrell) output-based measures of technical efficiency under
 assumption of CRS technology are computed for the following data:
    Number of data points (K) = 70
    Number of outputs     (M) = 3
    Number of inputs      (N) = 5
 Reference set is formed by 68 provided reference data points
. teradial y1 y2 y3 = x1 x2 x3 x4 x5, rts(nirs) base(output) reference(dref)
> tename(TErdNRSo) noprint
```

```
. teradial y1 y2 y3 = x1 x2 x3 x4 x5, rts(vrs) base(output) reference(dref)
> tename(TErdVRSo) noprint
. tenonradial y1 y2 y3 = x1 x2 x3 x4 x5, rts(crs) base(output) reference(dref)
> tename(TEnrCRSo) noprint
. tenonradial y1 y2 y3 = x1 x2 x3 x4 x5, rts(nirs) base(output) reference(dref)
> tename(TEnrNRSo) noprint
. tenonradial y1 y2 y3 = x1 x2 x3 x4 x5, rts(vrs) base(output) reference(dref)
> tename(TEnrVRSo) noprint
. list TErdCRSo TErdNRSo TErdVRSo TEnrCRSo TEnrNRSo TEnrVRSo in 1/7
```

	TErdCRSo	TErdNRSo	TErdVRSo	TEnrCRSo	TEnrNRSo	TEnrVRSo
1.	1.087257	1.032294	1.032294	1.11721	1.05654	1.05654
2.	1.110133	1.109314	1.109314	1.383089	1.277123	1.277123
3.	1.079034	1.068429	1.068429	1.17053	1.116582	1.116582
4.	1.119434	1.107413	1.107413	1.489086	1.471301	1.471301
5.	1.075864	1.075864	1	1.196779	1.196779	1
6.	1.107752	1.107752	1.105075	1.380214	1.378378	1.378378
7.	1.125782	1.119087	1.119087	1.575288	1.547186	1.547186

`teradial` and `tenonradial` compute measures of technical efficiency for all 70 data points using the reference set based on the restriction `x5 != 10`, which leaves two data points out. As expected, the radial measures are at least not worse than the nonradial measures for each returns to scale assumption. Figure 3 visualizes this observation, indicating slacks in outputs. However, for radial and nonradial measures, the measures under VRS are at least not worse than those under NIRS. The measures under NIRS are at least not worse than those under CRS. For data points 1, 2, 3, 4, and 7, measures under NIRS and VRS are equal. For data points 5 and 6, measures under NIRS and CRS are equal. We come back to scale analysis shortly when we discuss `nptestrts`.

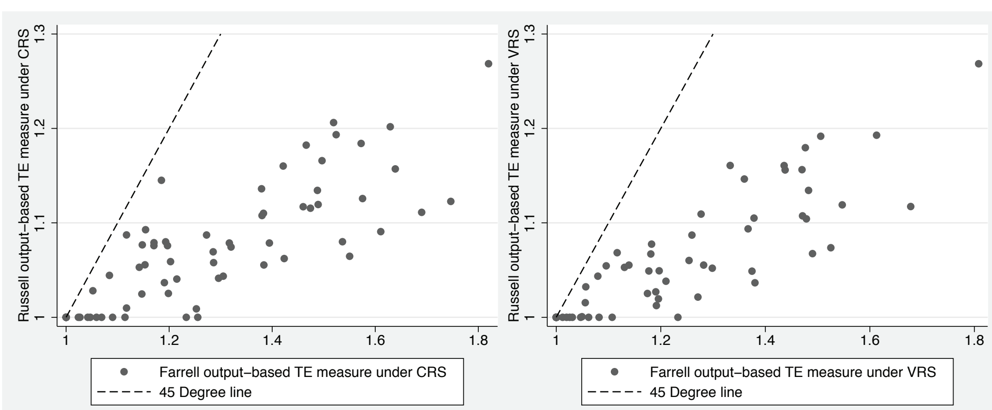

Figure 3. Scatterplot of Debreu–Farrell and RM of technical efficiency under the assumption of CRS (left panel) and under the assumption of VRS (right panel)

Before `teradialbc` is run, we need to know what type of bootstrap to use. We therefore perform the nonparametric test of independence by typing the new command `nptestind`. To illustrate, we run the test for all returns to scale assumptions for both output- and input-based frontier models. We show the output only for the output-based model under the assumption of CRS technology. We suppress the log for the remaining five models.

```
. matrix testsindpv = J(2, 3, .)
. matrix colnames testsindpv = CRS NiRS VRS
. matrix rownames testsindpv = output-based input-based
. nptestind y1 y2 y3 = x1 x2 x3 x4 x5, rts(crs) base(output) reps(999)
> alpha(0.05)
 Radial (Debreu-Farrell) output-based measures of technical efficiency
 under assumption of CRS, NIRS, and VRS technology are computed for the
 following data:
    Number of data points (K) = 70
    Number of outputs     (M) = 3
    Number of inputs      (N) = 5
 Reference set is formed by 70 data points, for which measures of
 technical efficiency are computed.
 ------------------------------------------------------------------------
         Test
 Ho: T4n = 0 (radial (Debreu-Farrell) output-based measure of technical
 efficiency under assumption of CRS technology and mix of outputs are
 independent)
 Bootstrapping test statistic T4n (999 replications)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
(dots omitted)
..................................................   950
.................................................
 p-value of the Ho that T4n = 0 (Ho that radial (Debreu-Farrell)
 output-based measure of technical efficiency under assumption of CRS
 technology and mix of outputs are independent) = 0.0671:
 hat{T4n} = 0.0310 is not statistically greater than 0 at the 5%
 significance level
 Heterogeneous bootstrap should be used when performing output-based
 technical efficiency measurement under assumption of CRS technology
. matrix testsindpv[1,1] = e(pvalue)
. nptestind y1 y2 y3 = x1 x2 x3 x4 x5, rts(nirs) base(output) reps(999)
> alpha(0.05) noprint
. matrix testsindpv[1,2] = e(pvalue)
. nptestind y1 y2 y3 = x1 x2 x3 x4 x5, rts(vrs) base(output) reps(999)
> alpha(0.05) noprint
. matrix testsindpv[1,3] = e(pvalue)
. nptestind y1 y2 y3 = x1 x2 x3 x4 x5, rts(crs) base(input) reps(999)
> alpha(0.05) noprint
. matrix testsindpv[2,1] = e(pvalue)
. nptestind y1 y2 y3 = x1 x2 x3 x4 x5, rts(nirs) base(input) reps(999)
> alpha(0.05) noprint
. matrix testsindpv[2,2] = e(pvalue)
```

```
. nptestind y1 y2 y3 = x1 x2 x3 x4 x5, rts(vrs) base(input) reps(999)
> alpha(0.05) noprint
. matrix testsindpv[2,3] = e(pvalue)
. matrix list testsindpv
testsindpv[2,3]
                    CRS       NiRS        VRS
output-based  .06706707   .2002002  .04204204
 input-based  .02902903    .003003  .22522523
```

Depending on the assumption about the technology and base of measurement, the `nptestind` command concludes differently about the type of bootstrap. In output-based efficiency measurement, the independence assumption is rejected at the 5% significance level only for VRS technology. In input-based efficiency measurement, it is only for VRS technology that the independence assumption is not rejected at the 5% significance level.

The `teradialbc` command can provide statistical inference for three types of bootstrap: 1) smoothed homogeneous, 2) smoothed heterogeneous, and 3) subsampling (heterogeneous).

We performed each of these types under the assumption of VRS technology. The results of `nptestind` indicate that the heterogeneous bootstrap should be used, so the results for the homogeneous bootstrap cannot be trusted. We report them here for illustrative purposes only. By using the `tebc()`, `biassqvar()`, `telower()`, and `teupper()` options, we generate new variables in the current dataset that contain bias-corrected output-based measures of technical efficiency, the statistic that compares the bias and the variance of the bootstrap (we also report its summary right after the command), and the lower and upper bounds of the 95% confidence interval for each of the three types of bootstrap. Table 1 lists the measures for the first 34 data points. We let `teradialbc` output the log and bootstrap dots for the first type but suppress them for the other two.

```
. teradialbc y1 y2 y3 = x1 x2 x3 x4 x5, rts(vrs) base(output) reference(dref)
> reps(999) tebc(TErdVRSoBC1) biassqvar(TErdVRSoBC1bv) telower(TErdVRSoLB1)
> teupper(TErdVRSoUB1)
 Radial (Debreu-Farrell) output-based measures of technical efficiency
 under assumption of VRS technology are computed for the following data:
    Number of data points (K) = 70
    Number of outputs     (M) = 3
    Number of inputs      (N) = 5
 Reference set is formed by 68 provided reference data points.
 ------------------------------------------------------------------------
 Bootstrapping reference set formed by 68 provided reference data points
 and computing radial (Debreu-Farrell) output-based measures of technical
 efficiency under assumption of VRS technology for each of 70 data points
 relative to the bootstrapped reference set
 Smoothed homogeneous bootstrap (999 replications)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
(dots omitted)
.................................................
```

```
. summarize TErdVRSoBC1bv

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
TErdVRSoBC~v |         70    3.648973    1.266609   1.927135   8.160317

. teradialbc y1 y2 y3 = x1 x2 x3 x4 x5, rts(vrs) base(output) reference(dref)
> heterogeneous reps(999) tebc(TErdVRSoBC2) biassqvar(TErdVRSoBC2bv)
> telower(TErdVRSoLB2) teupper(TErdVRSoUB2) noprint

. summarize TErdVRSoBC2bv

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
TErdVRSo~2bv |         58    40.70504    44.65367   5.851912   272.2176

. teradialbc y1 y2 y3 = x1 x2 x3 x4 x5, rts(vrs) base(output) reference(dref)
> subsampling reps(999) tebc(TErdVRSoBC3) biassqvar(TErdVRSoBC3bv)
> telower(TErdVRSoLB3) teupper(TErdVRSoUB3) noprint

. summarize TErdVRSoBC3bv

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
TErdVRSo~3bv |         70    3.428358    2.247708          0   14.89988
```

Statistic BV in table 1 is three times the ratio of bias squared to the variance of the bootstrap values of the radial measures of technical efficiency. Bias correction and statistical inference should be performed only if this statistic is well above unity. For all three types of bootstrap, the measure is satisfactory. For the smoothed homogeneous and subsampling bootstraps, the values of BV are much smaller than those for the smoothed heterogeneous bootstrap. If the BV is small, the variance of the bootstrap values is relatively large, and the mean squared error of the bias-corrected estimate of the technical efficiency measure is much higher than that of the original measure. We also know from running **nptestind** that the results from the smoothed homogeneous bootstrap should not be trusted.

Panels “Smoothed heterogeneous” and “Subsampling” in table 1 show the results for the heterogeneous bootstrap. Statistic BV indicates that the subsampling bootstrap introduces much noise. The bias-corrected measure is estimated imprecisely, and results for the subsampling bootstrap should not be used. This leaves us with reliable statistical inference using the heterogeneous smoothed bootstrap. Here the BV statistic is well above unity.

Table 1. Statistical inference about the radial output-based measure of technical efficiency under the assumption of VRS

#	TE[a]	Smoothed homogeneous				Smoothed heterogeneous				Subsampling			
		BC[b]	BV[c]	LB[d]	UB[e]	BC	BV	LB	UB	BC	BV	LB	UB
1	1.032	1.056	2.473	1.033	1.135	1.177	61.184	1.117	1.295	1.152	3.479	1.040	1.407
2	1.109	1.125	3.797	1.110	1.159	1.220	26.897	1.155	1.335	1.186	4.107	1.116	1.346
3	1.068	1.087	2.846	1.069	1.138	1.157	45.492	1.112	1.219	1.138	5.293	1.070	1.249
4	1.107	1.116	4.455	1.108	1.136	1.143	20.759	1.122	1.178	1.136	6.394	1.109	1.176
5	1.000	1.051	2.782	1.000	1.196	.	.	.	.	1.000	14.900	1.000	1.000
6	1.105	1.123	3.934	1.106	1.160	1.328	19.102	1.205	1.833	1.195	1.413	1.114	1.919
7	1.119	1.130	4.771	1.120	1.146	1.155	22.890	1.134	1.191	1.150	6.664	1.121	1.184
8	1.104	1.125	2.082	1.105	1.207	1.359	44.379	1.235	1.659	1.258	2.886	1.107	1.667
9	1.161	1.174	4.261	1.161	1.201	1.202	20.787	1.175	1.242	1.215	7.070	1.162	1.282
10	1.055	1.078	3.116	1.055	1.137	1.173	61.596	1.122	1.257	1.146	3.788	1.056	1.325
11	1.000	1.036	3.901	1.000	1.111	1.137	55.950	1.088	1.259	1.077	2.555	1.000	1.282
12	1.000	1.033	3.522	1.000	1.106	1.165	41.808	1.096	1.348	1.096	2.907	1.000	1.341
13	1.156	1.164	6.258	1.156	1.177	1.183	11.772	1.164	1.220	1.192	3.944	1.159	1.247
14	1.016	1.034	3.293	1.016	1.079	.	.	.	.	1.050	1.526	1.018	1.156
15	1.000	1.050	2.638	1.000	1.194	.	.	.	.	1.170	2.306	1.000	1.524
16	1.052	1.069	2.029	1.053	1.128	1.242	30.769	1.130	1.468	1.182	2.312	1.052	1.494
17	1.000	1.047	2.457	1.000	1.196	.	.	.	.	1.227	3.124	1.000	1.657
18	1.000	1.031	4.134	1.000	1.091	1.108	28.727	1.052	1.225	1.065	2.505	1.000	1.236
19	1.049	1.065	2.794	1.049	1.110	1.123	21.704	1.075	1.200	1.122	2.791	1.050	1.297
20	1.000	1.048	3.125	1.000	1.160	1.290	62.403	1.198	1.737	1.141	2.731	1.000	1.532
21	1.000	1.042	3.372	1.000	1.134	1.170	54.187	1.103	1.330	1.107	2.867	1.000	1.356
22	1.000	1.020	5.006	1.000	1.053	1.205	20.126	1.092	1.672	1.065	1.319	1.000	1.354
23	1.025	1.037	2.824	1.026	1.068	1.069	13.484	1.035	1.123	1.095	2.043	1.027	1.269
24	1.000	1.043	2.965	1.000	1.158	1.355	43.519	1.232	2.241	1.141	2.479	1.000	1.952
25	1.021	1.030	8.160	1.022	1.041	1.047	15.717	1.031	1.075	1.044	4.358	1.028	1.080
26	1.060	1.069	4.685	1.061	1.089	1.095	12.386	1.069	1.139	1.103	3.181	1.062	1.201
27	1.000	1.036	3.633	1.000	1.107	1.208	53.891	1.148	1.482	1.100	1.917	1.000	1.412
28	1.012	1.027	3.366	1.013	1.067	1.212	18.807	1.100	1.680	1.089	1.170	1.022	1.839
29	1.180	1.201	2.812	1.180	1.257	.	.	.	.	1.271	6.202	1.184	1.385
30	1.117	1.130	4.322	1.118	1.156	1.252	19.150	1.177	1.437	1.177	1.654	1.121	1.365
31	1.193	1.204	5.712	1.193	1.224	1.245	15.942	1.212	1.311	1.226	6.794	1.195	1.274
32	1.000	1.047	2.587	1.000	1.191	.	.	.	.	1.000	0.742	1.000	1.000
33	1.049	1.069	6.165	1.051	1.102	1.097	11.647	1.060	1.161	1.139	3.084	1.051	1.333
34	1.161	1.173	3.258	1.161	1.205	1.206	15.450	1.172	1.256	1.231	3.231	1.161	1.381

[a] original output-based measures of technical efficiency under assumption of VRS technology
[b] bias-corrected radial measures of technical efficiency
[c] three times the ratio of bias squared to variance for radial measures of technical efficiency
[d] lower-bound estimate for radial measures of technical efficiency
[e] upper-bound estimate for radial measures of technical efficiency

The bias for the heterogeneous smoothed bootstrap is larger than that of the homogeneous smoothed bootstrap, which is implied by larger bias-corrected estimates of efficiency measures. This means that the homogeneous bootstrap provides optimistic estimates of the bootstrapped frontier. The 95% confidence interval is also wider but not as wide as that for the subsampling bootstrap; this might be a result of the large variance of the bootstrap values for the subsampling bootstrap. Statistical inference cannot be provided for observations 5, 14, 15, 17, 29, and 32. The reason for this is too few bootstrap replications, where these observations lie within the bootstrap frontier, making the solution of the linear programming problem infeasible. Indeed, `e(realreps)` after `teradialbc` for the mentioned data points is 0(5), 1(14), 3(15), 0(17), 3(29), and 0(32).

Finally, we turn our discussion to the new command `nptestrts`, which performs nonparametric tests of returns to scale and analysis of scale efficiency. We have already determined that the heterogeneous smoothed bootstrap should be used for this dataset. We still provide the results using the homogeneous bootstrap and emphasize the caveat of using an incorrect bootstrap procedure.

```
. nptestrts y1 y2 y3 = x1 x2 x3 x4 x5, base(output) reps(999) alpha(0.05)
> testtwo sefficiency(SE_o) sefficient(SEffnt_hom) sineffdrs(SiDRS_hom)
 Radial (Debreu-Farrell) output-based measures of technical efficiency
 under assumption of CRS, NIRS, and VRS technology are computed for the
 following data:
    Number of data points (K) = 70
    Number of outputs     (M) = 3
    Number of inputs      (N) = 5
 Reference set is formed by 70 data points, for which measures of
 technical efficiency are computed.
 ------------------------------------------------------------------------
         Test #1
 Ho: mean(F_i^CRS)/mean(F_i^VRS) = 1
   and
 Ho: F_i^CRS/F_i^VRS = 1 for each of 70 data point(s)
 Bootstrapping reference set formed by 70 data points and computing
 radial (Debreu-Farrell) output-based measures of technical efficiency
 under assumption of CRS and VRS technology for each of 70 data points
 relative to the bootstrapped reference set
 Smoothed homogeneous bootstrap (999 replications)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
(dots omitted)
.................................................
 p-value of the Ho that mean(F_i^CRS)/mean(F_i^VRS) = 1 (Ho that the
 global technology is CRS) = 0.0010:
 mean(hat{F_i^CRS})/mean(hat{F_i^VRS}) = 1.0164 is statistically greater
 than 1 at the 5% significance level
 All data points are scale efficient
 ------------------------------------------------------------------------
```

```
        Test #2
 Ho: mean(F_i^NiRS)/mean(F_i^VRS) = 1
 Bootstrapping reference set formed by 70 data points and computing
 radial (Debreu-Farrell) output-based measures of technical efficiency
 under assumption of NIRS and VRS technology for each of 70 data points
 relative to the bootstrapped reference set
 Smoothed homogeneous bootstrap (999 replications)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
(dots omitted)
.................................................
 p-value of the Ho that mean(F_i^NiRS)/mean(F_i^VRS) = 1 (Ho that the
 global technology is NiRS) = 0.0030:
 mean(hat{F_i^NiRS})/mean(hat{F_i^VRS}) = 1.0085 is statistically greater
 than 1 at the 5% significance level

. table SEffnt_hom
```

Indicator variable if statistically scale efficient	Freq.
scale efficient	70

The *p*-value of the null hypothesis that the global technology is CRS (test #1) using the homogeneous smoothed bootstrap is very small, implying that CRS is not an appropriate assumption. Further, the null hypothesis that the global technology is NIRS (test #2) is also rejected. Hence, the nonparametric test of returns to scale advises performing efficiency measurement under the assumption of VRS technology. Additionally, the message `All data points are scale efficient` implies that test #1 is not rejected for a single data point. That the global returns to scale is not CRS is at odds with the latter finding. We now perform the test using the heterogeneous smoothed bootstrap:

```
. nptestrts y1 y2 y3 = x1 x2 x3 x4 x5, base(output) heterogeneous reps(999)
> alpha(0.05) testtwo sefficient(SEffnt_het) sineffdrs(SiDRS_het)
 Radial (Debreu-Farrell) output-based measures of technical efficiency
 under assumption of CRS, NIRS, and VRS technology are computed for the
 following data:
    Number of data points (K) = 70
    Number of outputs     (M) = 3
    Number of inputs      (N) = 5
 Reference set is formed by 70 data points, for which measures of
 technical efficiency are computed.
```

```
        Test #1
Ho: mean(F_i^CRS)/mean(F_i^VRS) = 1
  and
Ho: F_i^CRS/F_i^VRS = 1 for each of 70 data point(s)
Bootstrapping reference set formed by 70 data points and computing
radial (Debreu-Farrell) output-based measures of technical efficiency
under assumption of CRS and VRS technology for each of 70 data points
relative to the bootstrapped reference set
Smoothed heterogeneous bootstrap (999 replications)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
(dots omitted)
.................................................
p-value of the Ho that mean(F_i^CRS)/mean(F_i^VRS) = 1 (Ho that the
global technology is CRS) = 1.0000:
mean(hat{F_i^CRS})/mean(hat{F_i^VRS}) = 1.0164 is not statistically
greater than 1 at the 5% significance level
------------------------------------------------------------------------
        Test #2
Ho: F_i^NiRS/F_i^VRS = 1 for each of 1 scale inefficient data point(s)
Bootstrapping reference set formed by 70 data points and computing
radial (Debreu-Farrell) output-based measures of technical efficiency
under assumption of NIRS and VRS technology for each of 1 data points
relative to the bootstrapped reference set
Smoothed heterogeneous bootstrap (999 replications)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
(dots omitted)
.................................................
. table SEffnt_het
```

Indicator variable if statistically scale efficient	Freq.
scale inefficient	1
scale efficient	69

```
. table SiDRS_het
```

Indicator variable if statistically scale inefficient due to DRS	Freq.
scale inefficient due to DRS	1

Using the heterogeneous smoothed bootstrap, the nonparametric test fails to reject the null hypothesis that the global technology is CRS. This means there is no need to test that the global technology is NIRS (test #2). Performing test $\#1_k$, $k = 1, \ldots, 70$, however, suggests that one of the 70 data points is scale inefficient. Because we specified the `testtwo` option, test $\#2_k$ is performed for this single data point to determine the nature of its scale inefficiency. Using the `sineffdrs()` option, we generated the indicator variable `SiDRS_het` equal to 1 if the data point is statistically scale inefficient due to DRS. The `table SiDRS_het` command identifies that the data point is scale inefficient due to operating under the DRS portion of the technology [such as a data point (x_j, y_j) in terms of figure 1]. Table 2 lists the original measures of technical efficiency under the assumption of CRS and VRS technology, the scale efficiency measure, the indicator variables if statistically scale efficient, and the nature of scale inefficiency. Consider data point 1: that it is statistically scale efficient means that 1.053 is not statistically larger than 1. Consider data point 5: that it is not statistically scale efficient using the heterogeneous bootstrap means that 1.076 is statistically larger than 1.

Table 2. Scale analysis

#	CRS[a]	VRS[b]	SE	Scale efficient[c] (homogeneous)	Scale efficient (heterogeneous)	Scale inefficient due to DRS[d] (heterogeneous)
1	1.087	1.032	1.053	scale efficient	scale efficient	.
2	1.110	1.109	1.001	scale efficient	scale efficient	.
3	1.079	1.068	1.010	scale efficient	scale efficient	.
4	1.119	1.107	1.011	scale efficient	scale efficient	.
5	1.076	1.000	1.076	scale efficient	scale inefficient	scale inefficient due to DRS
6	1.108	1.105	1.002	scale efficient	scale efficient	.
7	1.126	1.119	1.006	scale efficient	scale efficient	.
8	1.111	1.104	1.006	scale efficient	scale efficient	.
9	1.184	1.161	1.020	scale efficient	scale efficient	.
10	1.077	1.055	1.021	scale efficient	scale efficient	.
11	1.025	1.000	1.025	scale efficient	scale efficient	.
12	1.028	1.000	1.028	scale efficient	scale efficient	.
13	1.166	1.156	1.009	scale efficient	scale efficient	.
14	1.076	1.016	1.059	scale efficient	scale efficient	.
15	1.000	1.000	1.000	scale efficient	scale efficient	.
16	1.065	1.052	1.012	scale efficient	scale efficient	.
17	1.000	1.000	1.000	scale efficient	scale efficient	.
18	1.000	1.000	1.000	scale efficient	scale efficient	.
19	1.058	1.049	1.008	scale efficient	scale efficient	.
20	1.000	1.000	1.000	scale efficient	scale efficient	.
21	1.000	1.000	1.000	scale efficient	scale efficient	.
22	1.000	1.000	1.000	scale efficient	scale efficient	.
23	1.044	1.025	1.018	scale efficient	scale efficient	.
24	1.000	1.000	1.000	scale efficient	scale efficient	.
25	1.041	1.021	1.020	scale efficient	scale efficient	.
26	1.074	1.060	1.013	scale efficient	scale efficient	.
27	1.000	1.000	1.000	scale efficient	scale efficient	.
28	1.059	1.012	1.046	scale efficient	scale efficient	.
29	1.206	1.180	1.023	scale efficient	scale efficient	.
30	1.123	1.117	1.005	scale efficient	scale efficient	.
31	1.202	1.193	1.007	scale efficient	scale efficient	.
32	1.117	1.000	1.117	scale efficient	scale efficient	.
33	1.079	1.049	1.028	scale efficient	scale efficient	.
34	1.182	1.161	1.019	scale efficient	scale efficient	.

[a] measure of technical efficiency under the assumption of CRS
[b] measure of technical efficiency under the assumption of VRS
[c] statistically scale efficient
[d] statistically scale inefficient due to DRS

The nonparametric test of returns to scale concludes differently about the null hypothesis depending on the type of bootstrap. This is not surprising because it is a consequence of using an inconsistent bootstrap procedure. The results of the test are based on correct mimicking of the data-generating process. If this is not guaranteed, then the bootstrap procedure is inconsistent and results of the nonparametric test cannot be trusted. For these particular data and base of efficiency measurement, the smoothed heterogeneous bootstrap should be used.

8.2 Data: PWT5.6

For our second dataset, we use the Penn World Tables, which were used by Kumar and Russell (2002) among others. See Summers and Heston (1991) for more details on the dataset. The purpose of this short study is to construct a Malmquist productivity index (MPI) between 1965 and 1990, and to perform analysis of the productivity change by decomposing the MPI. MPI uses the output distance function, which is the reciprocal of the Debreu–Farrell measure of technical efficiency (Caves, Christensen, and Diewert 1982; Färe, Grosskopf, and Lovell 1994a).

The Malmquist output-based productivity index from time period b to time period c for data point k is given by

$$\text{MPI}_k^{o,bc} = \left\{ \frac{F^o(y_{k,b}, x_{k,b}, y_b, x_b|\text{CRS})}{F^o(y_{k,c}, x_{k,c}, y_b, x_b|\text{CRS})} \times \frac{F^o(y_{k,b}, x_{k,b}, y_c, x_c|\text{CRS})}{F^o(y_{k,c}, x_{k,c}, y_c, x_c|\text{CRS})} \right\}^{1/2}$$

This index may be decomposed as

$$\begin{aligned} \text{MPI}_k^{o,bc} = & \frac{F^o(y_{k,b}, x_{k,b}, y_b, x_b|\text{CRS})}{F^o(y_{k,c}, x_{k,c}, y_c, x_c|\text{CRS})} \\ & \times \left\{ \frac{F^o(y_{k,c}, x_{k,c}, y_c, x_c|\text{CRS})}{F^o(y_{k,c}, x_{k,c}, y_b, x_b|\text{CRS})} \frac{F^o(y_{k,b}, x_{k,b}, y_c, x_c|\text{CRS})}{F^o(y_{k,b}, x_{k,b}, y_c, x_c|\text{CRS})} \right\}^{1/2} \end{aligned} \tag{10}$$

where $F^o(y_{k,d}, x_{k,d}, y_a, x_a|\text{CRS})$ is the Debreu–Farrell measure calculated for data point k in time period d to the frontier formed by observations (y_a, x_a) under the assumption of CRS technology. The first term in (10) measures the contribution of the technical efficiency change to the productivity change. The second term in (10) measures the contribution of the technical change to the productivity change:

$$\text{MPI} = \text{EFF} \times \text{TECH}$$

If EFF > 1 (< 1 in input-based measurement), then the change in efficiency has positively contributed to the productivity change from time period b to time period c. The meaning of TECH is the following: TECH $> / = / < 1$ implies that technical progress/stagnation/regress has occurred between periods b and c.

The decomposition of the MPI in (10) can be extended. Calculating the Debreu–Farrell measure under VRS, MPI can be decomposed into three components attributable to 1) pure technical efficiency change (PEFF), 2) technological change (TECH), and 3) scale efficiency change (SEC) (Färe et al. 1994b). The decomposition of the output-based MPI from time-period b to time period c for data point k is given by

$$\begin{aligned}\mathrm{MPI}_k^{o,bc} = & \frac{F^o(y_{k,b}, x_{k,b}, y_b, x_b|\mathrm{VRS})}{F^o(y_{k,c}, x_{k,c}, y_c, x_c|\mathrm{VRS})} \\ & \times \left\{ \frac{F^o(y_{k,c}, x_{k,c}, y_c, x_c|\mathrm{CRS})}{F^o(y_{k,c}, x_{k,c}, y_b, x_b|\mathrm{CRS})} \frac{F^o(y_{k,b}, x_{k,b}, y_c, x_c|\mathrm{CRS})}{F^o(y_{k,b}, x_{k,b}, y_c, x_c|\mathrm{CRS})} \right\}^{1/2} \\ & \times \frac{S_k^o(y_{k,b}, x_{k,b})}{S_k^o(y_{k,c}, x_{k,c})} \end{aligned} \tag{11}$$

where S_k^o is scale efficiency defined in (7).

From the outset, it is not clear which of the decompositions, (10) or (11), should be used. We first perform the nonparametric test of returns to scale using the heterogeneous bootstrap:

```
. use pwt56, clear
. reshape wide y k l, i(nu country) j(year)
(note: j = 1965 1990)
Data                               long   ->   wide
-----------------------------------------------------------------------------
Number of obs.                      104   ->      52
Number of variables                   6   ->       8
j variable (2 values)              year   ->   (dropped)
xij variables:
                                      y   ->   y1965 y1990
                                      k   ->   k1965 k1990
                                      l   ->   l1965 l1990
-----------------------------------------------------------------------------
. nptestrts y1965 = k1965 l1965, base(output) heterogeneous reps(999) alpha(0.05)
 Radial (Debreu-Farrell) output-based measures of technical efficiency
 under assumption of CRS, NIRS, and VRS technology are computed for the
 following data:
    Number of data points (K) = 52
    Number of outputs     (M) = 1
    Number of inputs      (N) = 2
 Reference set is formed by 52 data points, for which measures of
 technical efficiency are computed.
 ------------------------------------------------------------------------
         Test #1
 Ho: mean(F_i^CRS)/mean(F_i^VRS) = 1
   and
 Ho: F_i^CRS/F_i^VRS = 1 for each of 52 data point(s)
 Bootstrapping reference set formed by 52 data points and computing
 radial (Debreu-Farrell) output-based measures of technical efficiency
 under assumption of CRS and VRS technology for each of 52 data points
 relative to the bootstrapped reference set
 Smoothed heterogeneous bootstrap (999 replications)
```

```
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
(dots omitted)
..................................................
 p-value of the Ho that mean(F_i^CRS)/mean(F_i^VRS) = 1 (Ho that the
 global technology is CRS) = 0.9860:
 mean(hat{F_i^CRS})/mean(hat{F_i^VRS}) = 1.1196 is not statistically
 greater than 1 at the 5% significance level
 All data points are scale efficient
```

The null hypothesis that the global technology is CRS cannot be rejected. And anyway, all the countries are scale efficient, so that the third component in (11) is essentially 1. We therefore proceed with decomposition (10). We calculate the required efficiency measures as follows:

```
. teradial y1965 = k1965 l1965 (y1965 = k1965 l1965), rts(crs) base(output)
> tename(F11) noprint
. teradial y1990 = k1990 l1990 (y1965 = k1965 l1965), rts(crs) base(output)
> tename(F21) noprint
. teradial y1965 = k1965 l1965 (y1990 = k1990 l1990), rts(crs) base(output)
> tename(F12) noprint
. teradial y1990 = k1990 l1990 (y1990 = k1990 l1990), rts(crs) base(output)
> tename(F22) noprint
. generate mpi = sqrt(F12 / F22 * F22 / F21)
(7 missing values generated)
. generate effch = F11 / F22
. generate techch = mpi / effch
(7 missing values generated)
```

We present the results of the decomposition for the first 34 out of 52 data points in table 3, and we discuss selected results.

Argentina was on the frontier in 1965 but moved away from the 1990 frontier. Hong Kong was quite inefficient in 1965, but in 1990 it defines the frontier. We also observe that the productivity of industrialized countries such as Australia, Austria, and Belgium has increased, while productivity has fallen for Argentina, Bolivia, and Ecuador, among others. The productivity has increased, for example, in Australia because of both improved efficiency and improved technology. In Bolivia, the main reason for decreased productivity was loss in efficiency. In Malawi, on the contrary, efficiency change has positively contributed to the growth of productivity, but technology has deteriorated so much that overall productivity has decreased.

Table 3. Measures of technical efficiency and MPI

#	Country	1965[a]	1990[b]	MPI	EFFch	TECHch
1	Argentina	1.000	1.546	0.818	0.647	1.264
2	Australia	1.277	1.202		1.062	
3	Austria	1.174	1.374	1.260	0.854	1.475
4	Belgium	1.419	1.153	1.549	1.231	1.258
5	Bolivia	2.002	2.457	0.948	0.815	1.163
6	Canada	1.247	1.046	.	1.193	.
7	Chile	1.180	1.549	0.889	0.762	1.167
8	Columbia	2.415	2.243	1.062	1.077	0.987
9	Denmark	1.324	1.432	1.240	0.924	1.342
10	Dominican Rep.	1.383	1.953	0.914	0.708	1.291
11	Equador	2.664	2.756	0.961	0.966	0.994
12	Finland	1.963	1.303	.	1.506	.
13	France	1.257	1.208	1.437	1.040	1.381
14	Germany, West	1.450	1.204	.	1.204	
15	Greece	1.828	1.673	1.119	1.093	1.025
16	Guatemala	1.228	1.369	1.037	0.897	1.156
17	Honduras	2.224	2.431	1.022	0.915	1.117
18	Hong Kong	2.202	1.000	1.519	2.202	0.690
19	Iceland	1.041	1.146	0.997	0.909	1.097
20	India	2.723	2.417	1.226	1.127	1.088
21	Ireland	1.411	1.184	1.134	1.192	0.952
22	Israel	1.664	1.192	1.238	1.396	0.887
23	Italy	1.490	1.131	1.443	1.318	1.096
24	Jamaica	1.774	1.930	1.017	0.919	1.107
25	Japan	1.684	1.617	1.447	1.041	1.390
26	Kenya	3.902	3.411	1.328	1.144	1.161
27	Korea, Rep	2.309	1.632	1.228	1.415	0.868
28	Malawi	3.515	2.996	0.621	1.173	0.529
29	Mauritius	1.062	1.025	1.115	1.036	1.076
30	Mexico	1.171	1.347	0.950	0.869	1.093
31	Netherlands	1.190	1.130	1.285	1.054	1.219
32	New Zealand	1.175	1.406	1.148	0.836	1.373
33	Norway	1.000	1.210	.	0.826	.
34	Panama	2.266	3.021	0.859	0.750	1.146

[a] measure of technical efficiency under the assumption of CRS in 1965
[b] measure of technical efficiency under the assumption of CRS in 1990

9 Sample restriction, discussion, and runtime

❏ **Technical note**

All functions create Stata matrices and feed them to plugins. The number of data points that can be used in all functions is thus limited by `matsize` (see [R] **matsize**). Stata/IC allows a maximum of 800, while Stata/MP and Stata/SE allow up to 11,000. ❏

❏ **Technical note**

Stata 11.2 and above can be used to run the new commands described in this article. Earlier versions of Stata can probably also be used but have not been tested. ❏

❏ **Technical note**

Plugins for solving linear programming problems use the quickhull algorithm (Barber, Dobkin, and Huhdanpaa [1996]; http://www.qhull.org/) and GLPK version 4.55 (GNU Linear Programming Kit [2012]; available at http://www.gnu.org/software/glpk/) coded in C. The required plugins are compiled with C code. The plugins are available for Mac OS X, Ubuntu, and Windows systems. ❏

Because the linear programming is coded in low-level language, the new commands are very fast. We have recorded the time required to do the calculations in this paper. The calculations were computed on an iMac (late 2012) desktop with a 2.9 GHz processor.

```
. timer clear
. timer on 1
. use ccr81, clear
(Program Follow Through at 70 US Primary Schools)
. generate dref = x5 != 10
. tenonradial y1 y2 y3 = x1 x2 x3 x4 x5, rts(crs) base(output) reference(dref)
> tename(TErdCRSo) noprint
. timer off 1
. timer on 2
. nptestind y1 y2 y3 = x1 x2 x3 x4 x5, rts(crs) base(output) reps(999)
> alpha(0.05) noprint
. timer off 2
. timer on 3
. teradialbc y1 y2 y3 = x1 x2 x3 x4 x5, rts(vrs) base(output) reference(dref)
> reps(999) tebc(TErdVRSoBC1) biassqvar(TErdVRSoBC1bv) telower(TErdVRSoLB1)
> teupper(TErdVRSoUB1) noprint
. timer off 3
. timer on 4
. teradialbc y1 y2 y3 = x1 x2 x3 x4 x5, rts(vrs) base(output) reference(dref)
> heterogeneous reps(999) tebc(TErdVRSoBC2) biassqvar(TErdVRSoBC2bv)
> telower(TErdVRSoLB2) teupper(TErdVRSoUB2) noprint
. timer off 4
. timer on 5
. nptestrts y1 y2 y3 = x1 x2 x3 x4 x5, base(output) reps(999) alpha(0.05)
> sefficient(SEffnt_hom)
 Radial (Debreu-Farrell) output-based measures of technical efficiency
 under assumption of CRS, NIRS, and VRS technology are computed for the
 following data:
    Number of data points (K) = 70
    Number of outputs     (M) = 3
    Number of inputs      (N) = 5
 Reference set is formed by 70 data points, for which measures of
 technical efficiency are computed.
 ---------------------------------------------------------------------
```

```
        Test #1
 Ho: mean(F_i^CRS)/mean(F_i^VRS) = 1
   and
 Ho: F_i^CRS/F_i^VRS = 1 for each of 70 data point(s)
 Bootstrapping reference set formed by 70 data points and computing
 radial (Debreu-Farrell) output-based measures of technical efficiency
 under assumption of CRS and VRS technology for each of 70 data points
 relative to the bootstrapped reference set
 Smoothed homogeneous bootstrap (999 replications)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100
..................................................   150
..................................................   200
..................................................   250
..................................................   300
..................................................   350
..................................................   400
..................................................   450
..................................................   500
..................................................   550
..................................................   600
..................................................   650
..................................................   700
..................................................   750
..................................................   800
..................................................   850
..................................................   900
..................................................   950
.................................................
 p-value of the Ho that mean(F_i^CRS)/mean(F_i^VRS) = 1 (Ho that the
 global technology is CRS) = 0.0050:
 mean(hat{F_i^CRS})/mean(hat{F_i^VRS}) = 1.0164 is statistically greater
 than 1 at the 5% significance level
 All data points are scale efficient
 ------------------------------------------------------------------------

. timer off 5
. timer on 6
. nptestrts y1 y2 y3 = x1 x2 x3 x4 x5, base(output) heterogeneous reps(999)
> alpha(0.05) sefficient(SEffnt_het)
 Radial (Debreu-Farrell) output-based measures of technical efficiency
 under assumption of CRS, NIRS, and VRS technology are computed for the
 following data:
    Number of data points (K) = 70
    Number of outputs     (M) = 3
    Number of inputs      (N) = 5
 Reference set is formed by 70 data points, for which measures of
 technical efficiency are computed.
 ------------------------------------------------------------------------
```

```
        Test #1
Ho: mean(F_i^CRS)/mean(F_i^VRS) = 1
  and
Ho: F_i^CRS/F_i^VRS = 1 for each of 70 data point(s)
Bootstrapping reference set formed by 70 data points and computing
radial (Debreu-Farrell) output-based measures of technical efficiency
under assumption of CRS and VRS technology for each of 70 data points
relative to the bootstrapped reference set
Smoothed heterogeneous bootstrap (999 replications)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100
..................................................   150
..................................................   200
..................................................   250
..................................................   300
..................................................   350
..................................................   400
..................................................   450
..................................................   500
..................................................   550
..................................................   600
..................................................   650
..................................................   700
..................................................   750
..................................................   800
..................................................   850
..................................................   900
..................................................   950
.................................................
p-value of the Ho that mean(F_i^CRS)/mean(F_i^VRS) = 1 (Ho that the
global technology is CRS) = 1.0000:
mean(hat{F_i^CRS})/mean(hat{F_i^VRS}) = 1.0164 is not statistically
greater than 1 at the 5% significance level
All data points are scale efficient
--------------------------------------------------------------------------

. timer off 6
. timer list
   1:      0.13 /        1 =       0.1330
   2:      2.63 /        1 =       2.6350
   3:      9.26 /        1 =       9.2560
   4:     19.85 /        1 =      19.8490
   5:    299.50 /        1 =     299.5040
   6:   1095.99 /        1 =    1095.9890
```

`tenonradial` is trivial and runs instantly. This remains true even if the dataset has a sample size of several thousands. `nptestind` is also quite fast but may slow down as the sample size grows. The smoothed homogeneous bootstrap in `teradialbc` runs relatively quickly on small samples (9 seconds), but the heterogeneous bootstrap is more demanding (20 seconds) and that time will increase with sample size. Running the nonparametric test of returns to scale, `nptestrts`, is the most involved because, instead of doing calculations for each of K data points on each bootstrap replication as in `teradialbc`, the binomial test requires bootstrap replications for each of K data points independently. This is time demanding, especially when the smoothed heterogeneous bootstrap is used. On a sample of 70 data points, it took about 5 minutes for the homogeneous bootstrap and 18 minutes for the heterogeneous bootstrap.

Displaying dots does not add to the output but rather indicates how long the whole bootstrap is going to take. It can be suppressed by specifying the `nodots` option in each of the new commands.

10 Comparison with the dea command

The new command `teradial` performs radial technical efficiency analysis, which the user-written command `dea` (Ji and Lee 2010) also offers. The latter command has two serious limitations for a practitioner. First, it is slow with even moderate datasets. We have recorded the time it takes to compute an input-based measure of technical efficiency under VRS using both commands for samples of 10 to 70 in steps of 10 data points:

```
. use ccr81, clear
(Program Follow Through at 70 US Primary Schools)
. rename nu dmu
. timer clear
. * number of observations: 10(10)70
. forvalues nobs = 10(10)70{
  2.  local nobs = `nobs´
  3.  local nobs2 = `nobs´ + 1
  4.  timer on `nobs´
  5.  quietly dea x1 x2 x3 x4 x5 = y1 y2 y3 in 1/`nobs´, rts(vrs) ort(in)
  6.  timer off `nobs´
  7.  timer on `nobs2´
  8.  quietly teradial y1 y2 y3 = x1 x2 x3 x4 x5 in 1/`nobs´, rts(vrs)
> base(input) tename(TErdVRSi_`nobs´)
  9.  timer off `nobs2´
 10. }
. timer list
  10:      8.85 /        1 =       8.8470
  11:      0.00 /        1 =       0.0030
  20:     33.87 /        1 =      33.8680
  21:      0.01 /        1 =       0.0070
  30:     71.71 /        1 =      71.7090
  31:      0.02 /        1 =       0.0180
  40:    849.43 /        1 =     849.4260
  41:      0.04 /        1 =       0.0360
  50:   1067.00 /        1 =    1066.9960
  51:      0.09 /        1 =       0.0890
  60:   1839.14 /        1 =    1839.1440
  61:      0.15 /        1 =       0.1470
  70:   1990.05 /        1 =    1990.0520
  71:      0.22 /        1 =       0.2180
```

The time it takes increases for both `dea` and `teradial`, but `dea` becomes slow very quickly. For a sample size of 70 data points, `teradial` completes in a fifth of a second, while `dea` needs 33 minutes. This can be a real bottleneck in actual empirical analysis. Thus, using `dea` for making statistical inference is next to infeasible.

Second, teradial can calculate the measure of technical efficiency of a data point relative to the frontier defined by the user by specifying the reference() option. This is required, for example, for analysis of productivity change as demonstrated in section 8.2. Such analysis is not possible using dea.

11 Concluding remarks

We introduced five new Stata commands that fit and provide statistical inference in nonparametric frontier models. tenonradial and teradial calculate nonradial Russell and radial Debreu–Farrell measures of technical efficiency, respectively. The measures can be computed for different assumptions about the technology and base of the analysis, as well as relative to the frontier formed by data points provided by the user. These frontier models are deterministic, and resulting measures are subject to sampling variation.

teradialbc can accommodate different types of bootstrapping techniques to provide statistical inference regarding these deterministic measures. For obtaining reliable results from teradialbc, the bootstrap type must correctly mimic the data-generating process. nptestind provides a simple tool to determine the type of bootstrap consistent with the data. Finally, nptestrts uses the bootstrap to provide inference with regard to the underlying technology and performs a scale analysis of each data point.

We have presented two empirical examples. In the first example, we illustrated the capabilities of the new commands and discussed the implications for empirical analysis. In the second example, we showed that these commands can be used to analyze the changes in productivity for 52 countries from 1965 to 1990.

12 Acknowledgments

The authors are grateful to the Lebesgue Center of Mathematics, which supported the work of Pavlo Mozharovskyi according to the program PIA-ANR-11-LABX-0020-01. We also thank H. Joseph Newton and an anonymous reviewer for comments.

13 References

Badunenko, O., D. J. Henderson, and S. C. Kumbhakar. 2012. When, where and how to perform efficiency estimation. *Journal of the Royal Statistical Society, Series A* 175: 863–892.

Barber, C. B., D. P. Dobkin, and H. Huhdanpaa. 1996. The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software* 22: 469–483.

Caves, D. W., L. R. Christensen, and W. E. Diewert. 1982. The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica* 50: 1393–1414.

Charnes, A., W. W. Cooper, and E. Rhodes. 1981. Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through. *Management Science* 27: 668–697.

Cobb, C. W., and P. H. Douglas. 1928. A theory of production. *American Economic Review* 18 (Suppl. 1): 139–165.

Coelli, T. J., D. S. P. Rao, C. J. O'Donnell, and G. E. Battese. 2005. *An Introduction to Efficiency and Productivity Analysis*. 2nd ed. New York: Springer.

Debreu, G. 1951. The coefficient of resource utilization. *Econometrica* 19: 273–292.

Färe, R. 1998. *Fundamentals of Production Theory*. Berlin: Springer.

Färe, R., and S. Grosskopf. 1985. A nonparametric cost approach to scale efficiency. *Scandinavian Journal of Economics* 87: 594–604.

Färe, R., S. Grosskopf, and C. A. K. Lovell. 1994a. *Production Frontiers*. Cambridge: Cambridge University Press.

Färe, R., S. Grosskopf, M. Norris, and Z. Zhang. 1994b. Productivity growth, technical progress, and efficiency change in industrialized countries. *American Economic Review* 84: 66–83.

Färe, R., and C. A. K. Lovell. 1978. Measuring the technical efficiency of production. *Journal of Economic Theory* 19: 150–162.

Färe, R., and D. Primont. 1995. *Multi-Output Production and Duality: Theory and Applications*. New York: Kluwer.

Farrell, M. J. 1957. The measurement of productive efficiency. *Journal of the Royal Statistical Society, Series A* 120: 253–290.

GNU Linear Programming Kit. 2012. http://www.gnu.org/software/glpk/glpk.html.

Ji, Y.-B., and C. Lee. 2010. Data envelopment analysis. *Stata Journal* 10: 267–280.

Kneip, A., L. Simar, and P. W. Wilson. 2008. Asymptotics and consistent bootstraps for DEA estimators in nonparametric frontier models. *Econometric Theory* 24: 1663–1697.

Koopmans, T. C. 1951. An analysis of production as an efficient combination of activities. In *Activity Analysis of Production and Allocation*, ed. T. C. Koopmans. New York: Wiley.

Kumar, S., and R. R. Russell. 2002. Technological change, technological catch-up, and capital deepening: Relative contributions to growth and convergence. *American Economic Review* 92: 527–548.

Simar, L., and P. W. Wilson. 1998. Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science* 44: 49–61.

———. 2000. A general methodology for bootstrapping in non-parametric frontier models. *Journal of Applied Statistics* 27: 779–802.

———. 2002. Non-parametric tests of returns to scale. *European Journal of Operational Research* 139: 115–132.

Summers, R., and A. Heston. 1991. The Penn World Table (Mark 5): An expanded set of international comparisons, 1950–1988. *Journal of the Royal Statistical Society, Series A* 106: 327–368.

Wilson, P. W. 2003. Testing independence in models of productive efficiency. *Journal of Productivity Analysis* 20: 361–390.

About the authors

Oleg Badunenko is an assistant professor at the University of Cologne in Germany. His major research interests are efficiency and productivity analysis, as well as micro- and nonparametric econometrics.

Pavlo Mozharovskyi is a postdoc researcher at the Centre Henri Lebesgue, Institute of Mathematical Research of Rennes and Agrocampus Ouest in Rennes, France. His main research interests lie in nonparametric and computational statistics, classification, and imputation of missing data.