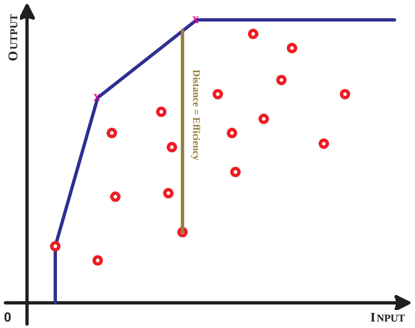
Nonparametric Frontier Analysis using Stata



Oleg Badunenko

University of Cologne

Pavlo Mozharovskyi

Centre Henri Lebesgue,
Agrocampus Ouest, Rennes

2016 German Stata Users Group Meeting

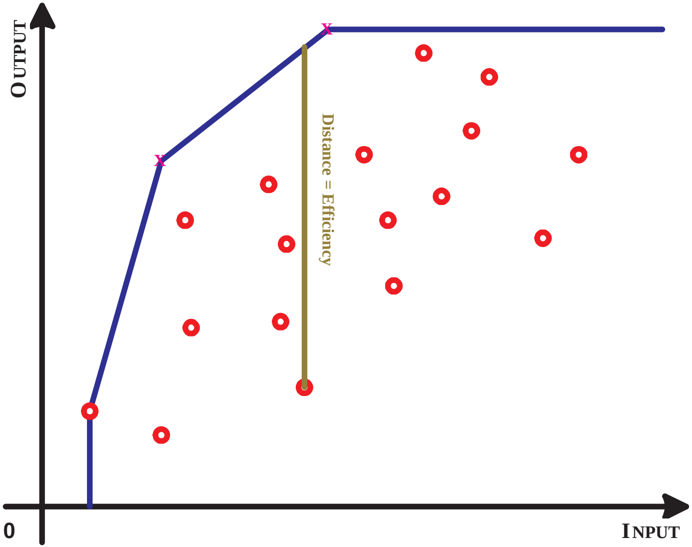
Overview

1. Nonparametric frontier analysis
   - Motivation
   - Radial efficiency analysis
   - Nonradial efficiency analysis
2. Statistical inference in radial frontier model
   - Type of the bootstrap for statistical inference
   - Returns to scale and scale analysis
3. Stata commands
   - `tenonradial`, `teradial`, `teradialbc`, `nptestind`, and `nptestrts`
4. Empirical application
   - Data: CCR81
   - Data: PWT5.6
5. Discussion
   - Sample restriction and runtime
   - Comparison to `dea` command in Stata, Stata Journal, 10(2): 267-80
   - Concluding remarks

Outline

1. Nonparametric frontier analysis
   - Motivation
   - Radial efficiency analysis
   - Nonradial efficiency analysis
2. Statistical inference in radial frontier model
   - Type of the bootstrap for statistical inference
   - Returns to scale and scale analysis
3. Stata commands
   - `tenonradial`, `teradial`, `teradialbc`, `nptestind`, and `nptestrts`
4. Empirical application
   - Data: CCR81
   - Data: PWT5.6
5. Discussion
   - Sample restriction and runtime
   - Comparison to dea command in Stata, Stata Journal, 10(2): 267-80
   - Concluding remarks

Technical efficiency measurement



Radial efficiency analysis

- 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.

Radial efficiency analysis (cont.)

- 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 a measure of technical efficiency.
  - In input-based radial efficiency measurement, it is the amount of necessary (proportional) reduction of inputs to move a data point to a boundary of the input requirement set $L(y)$.

Radial efficiency analysis (cont.)

- Hypothetical one-input one-output production processes with three different technologies CRS, VRS and NIRS

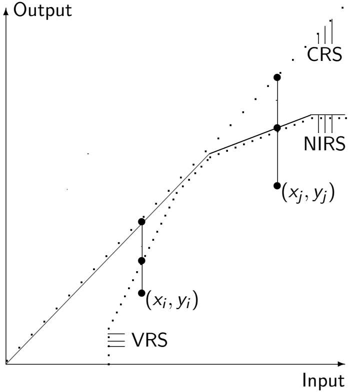


Figure: Output-based

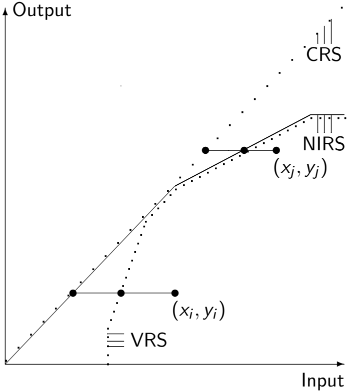


Figure: Input-based

Radial efficiency analysis (cont.)

- Empirically, 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 ($k = 1, \ldots, K$):

$$\hat{F}_k^o(y_k, x_k, y, x | \text{CRS}) = \max_{\theta, z} \theta$$

$$\text{s.t.} \sum_{k=1}^{K} z_k y_{km} \geq y_{km} \theta_m, m = 1, \cdots, M,$$

$$\sum_{k=1}^{K} z_k x_{kn} \leq x_{kn}, n = 1, \cdots, N,$$

$$z_k \geq 0.$$

y is $K \times M$ matrix of available data on outputs, x is $K \times N$ matrix of available data on inputs.

Radial efficiency analysis (cont.)

- This specifies constant returns to scale technology (CRS).
- For variable returns to scale (VRS) a convexity constraint $\sum_{k=1}^{K} z_k = 1$ is added, while
- for non-increasing returns to scale (NIRS), $\sum_{k=1}^{K} z_k \leq 1$ inequality is added.

Nonradial efficiency analysis

- For data point (y_k, x_k), radial measure expands (shrinks) all M outputs $y_k = (y_{k1}, \ldots, y_{kM})$ (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.
- Nonradial measure of technical efficiency, the Russell measure.

Nonradial efficiency analysis (cont.)

- Output based,

$$RM_k^o(y_k, x_k, y, x|\text{CRS}) = \max\left\{ \begin{array}{r} M^{-1}\sum_{m=1}^{M}\theta_m\colon\ (\theta_1 y_{k1}, \dots \theta_M y_{kM}) \in P(x), \\ \theta_m \geq 0, m = 1, \dots, M \end{array} \right\}. \quad (4)$$

- The input-based counterpart is given by

$$RM_k^i(y_k, x_k, y, x|\text{CRS}) = \min\left\{ \begin{array}{r} N^{-1}\sum_{n=1}^{N}\lambda_n\colon\ (\lambda_1 x_{k1}, \dots \lambda_N y_{kN}) \in L(y), \\ \lambda_n \geq 0, n = 1, \dots, N \end{array} \right\}. \quad (5)$$

Nonradial efficiency analysis (cont.)

- Since the Russell measure can expand (shrink) an output (input) vector at most (least) as far as the radial measure can, we have the result that

$$1 \geq \widehat{RM}_k^o(y_k, x_k, y, x|\mathrm{CRS}) \geq \hat{F}_k^o(y_k, x_k, y, x|\mathrm{CRS}) \tag{6}$$

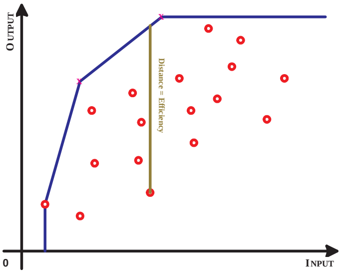
and

$$0 < \widehat{RM}_k^i(y_k, x_k, y, x|\mathrm{CRS}) \leq \hat{F}_k^i(y_k, x_k, y, x|\mathrm{CRS}) \leq 1. \tag{7}$$

Outline

1. Nonparametric frontier analysis
   - Motivation
   - Radial efficiency analysis
   - Nonradial efficiency analysis
2. Statistical inference in radial frontier model
   - Type of the bootstrap for statistical inference
   - Returns to scale and scale analysis
3. Stata commands
   - `tenonradial`, `teradial`, `teradialbc`, `nptestind`, and `nptestrts`
4. Empirical application
   - Data: CCR81
   - Data: PWT5.6
5. Discussion
   - Sample restriction and runtime
   - Comparison to dea command in Stata, Stata Journal, 10(2): 267-80
   - Concluding remarks

Statistical inference



- The estimated technical efficiency measures are *too optimistic*, due to the fact that the DEA estimate of the production set is necessarily 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 bootstrap technique.

Type of the bootstrap for statistical inference

- The major assumption of the bootstrapping technique depends on whether the estimated output-based measures of technical efficiency are independent of the mix of outputs.
- This dependency is testable given the assumption of returns to scale of the global technology (`nptestind` command).
- If output-based measures of technical efficiency are independent of the mix of outputs, the smoothed homogeneous bootstrap can be used.
- This type of the bootstrap is not computer intensive.
- Otherwise, the heterogenous bootstrap must be used to provide valid statistical inference.
- The latter type of bootstrap is quite computer demanding and may take a while for large data sets.
- Both implemented in `teradialbc`.

Technology and scale analysis

- The assumption regarding the global technology is crucial in DEA.
- 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, the returns to scale assumption can be tested econometrically.
- Moreover, if technology is not CRS globally, estimating measure of technical efficiency under CRS will lead to inconsistent results.

Technology and scale analysis (cont.)

- The following tests are implemented in `nptestrts`

$$\text{Test } \#1: \quad H_0 : T \text{ is globally CRS}$$
$$H_1 : T \text{ is VRS.}$$

- If null hypothesis H_0 is rejected, that is, technology is not CRS everywhere, the following test with less restrictive null hypothesis may be performed

$$\text{Test } \#2: \quad H_0' : T \text{ is globally NIRS}$$
$$H_1 : T \text{ is VRS.}$$

Technology and scale analysis (cont.)

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

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

and

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

Technology and scale analysis (cont.)

- Additionally, this testing procedure can be used to perform the scale analysis for each data point.

Outline

1. Nonparametric frontier analysis
   - Motivation
   - Radial efficiency analysis
   - Nonradial efficiency analysis
2. Statistical inference in radial frontier model
   - Type of the bootstrap for statistical inference
   - Returns to scale and scale analysis
3. Stata commands
   - `tenonradial`, `teradial`, `teradialbc`, `nptestind`, and `nptestrts`
4. Empirical application
   - Data: CCR81
   - Data: PWT5.6
5. Discussion
   - Sample restriction and runtime
   - Comparison to dea command in Stata, Stata Journal, 10(2): 267-80
   - Concluding remarks

tenonradial

tenonradial uses reduced linear programming to compute the nonradial output- or input-based measure of technical efficiency, which is known as the Russell measure.

tenonradial *outputs* = *inputs* [(*ref_outputs* = *ref_inputs*)] [*if*] [*in*] [, rts(*string*) base(*string*) ref(*varname*) tename(*newvarname*) noprint]

- *outputs* is the list of output variables
- *inputs* is the list of input variables
- rts(*rtsassumption)* specifies returns to scale assumption
- base(*basetype)* specifies type of optimization
- tenames(*newvarname*) creates *newvarname* containing the nonradial measures of technical efficiency.

teradial

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

teradialbc

teradialbc performs statistical inference about the radial measure of technical efficiency

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

Bootstrap options: subsampling, kappa(#) , smoothed, heterogeneous, reps(#), level(#)
Variable generation: tenames(*newvarname*), tebc(*newvarname*), biasboot(*newvarname*), varboot(*newvarname*), biassqvar(*newvarname*), telower(*newvarname*), teupper(*newvarname*)

nptestind

nptestind performs nonparametric test of independence

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

nptestrts

nptestrts performs nonparametric test of returns to scale

nptestrts *outputs* = *inputs* [*if*] [*in*] [, rts(*string*) base(*string*)
ref(*varname*) heterogeneous reps(#) alpha(#) testtwo
tecrsname(*newvarname*) tenrsname(*newvarname*) tevrsname(*newvarname*)
sefficiency(*newvarname*) psefficient(*newvarname*)
sefficient(*newvarname*) nrsovervrs(*newvarname*)
pineffdrs(*newvarname*) sineffdrs(*newvarname*) noprint nodots]

Testing: testtwo specifies that the Test #2 is performed.
Variable generation: related to testing procedure

Outline

1. Nonparametric frontier analysis
   - Motivation
   - Radial efficiency analysis
   - Nonradial efficiency analysis
2. Statistical inference in radial frontier model
   - Type of the bootstrap for statistical inference
   - Returns to scale and scale analysis
3. Stata commands
   - `tenonradial`, `teradial`, `teradialbc`, `nptestind`, and `nptestrts`
4. Empirical application
   - Data: CCR81
   - Data: PWT5.6
5. Discussion
   - Sample restriction and runtime
   - Comparison to dea command in Stata, Stata Journal, 10(2): 267-80
   - Concluding remarks

Data: CCR81

- Originally used to evaluate the efficiency of public programs and their management
  - 5 inputs
  - 3 outputs
- Artificially create a variable `dref` to illustrate the capabilities of new commands

CCR81: Technical efficiency

```
. use ccr81, clear
(Program Follow Through at 70 US Primary Schools)

. generate dref = x5 != 10 /// (2 such data points)

. teradial y1 y2 y3 = x1 x2 x3 x4 x5, r(c) b(o) ref(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, r(n) b(o) ref(dref) tename(TErdNRSo) nopr
> int

. teradial y1 y2 y3 = x1 x2 x3 x4 x5, r(v) b(o) ref(dref) tename(TErdVRSo) nopr
> int

. tenonradial y1 y2 y3 = x1 x2 x3 x4 x5, r(c) b(o) ref(dref) tename(TEnrCRSo) n
> oprint

. tenonradial y1 y2 y3 = x1 x2 x3 x4 x5, r(n) b(o) ref(dref) tename(TEnrNRSo) n
> oprint

. tenonradial y1 y2 y3 = x1 x2 x3 x4 x5, r(v) b(o) ref(dref) tename(TEnrVRSo) n
> oprint
```

CCR81: Technical efficiency (cont.)

```
. 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

CCR81: Technical efficiency (cont.)

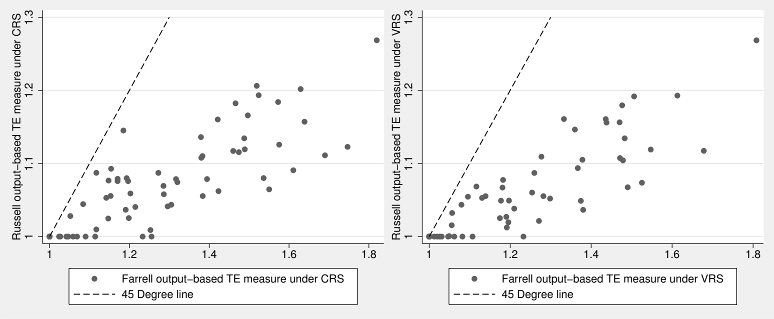


Figure: Scatterplot of Debreu-Farrell and Russell measures of technical efficiency under the assumption of CRS (left panel) and under the assumption of VRS (right panel)

CCR81: Test of independence

```
. 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, r(c) b(o) reps(999) a(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)

 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) = 0.0621:

 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)
fill the rest of the matrix
. matrix list testsindpv

testsindpv[2,3]
                    CRS        NiRS         VRS
output-based  .06206206   .21921922   .03803804
 input-based  .02402402   .00500501   .24624625
```

CCR81: Bias-correction. Should it be used?

```
. teradialbc y1 y2 y3 = x1 x2 x3 x4 x5, r(v) b(o) ref(dref) reps(999) tebc(TErd
> VRSoBC1) 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)
 ------------------------------------------------------------------------------

 Warning: for one or more data points statistic [(1/3)*(bias)^2/var] is smaller
 than 1; bias-correction should not be used

. summarize TErdVRSoBC1bv

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
TErdVRSoBC~v |        70    .4082286    .1514485   .2270871   .9380344
. teradialbc y1 y2 y3 = x1 x2 x3 x4 x5, het r(v) b(o) ref(dref) reps(999) tebc(
> TErdVRSoBC2) biassqvar(TErdVRSoBC2bv) telower(TErdVRSoLB2) teupper(TErdVRSoUB
> 2) noprint

. summarize TErdVRSoBC2bv

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
TErdVRSo~2bv |        58    4.557299    4.873043   .7694204   27.62238
. teradialbc y1 y2 y3 = x1 x2 x3 x4 x5, subs r(v) b(o) ref(dref) reps(999) tebc
> (TErdVRSoBC3) biassqvar(TErdVRSoBC3bv) telower(TErdVRSoLB3) teupper(TErdVRSoU
> B3) noprint

. summarize TErdVRSoBC3bv

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
TErdVRSo~3bv |        70    .3868087    .2562643          0     1.6995
```

CCR81: Nonparametric test of returns to scale: it's CRS

```
. nptestrts y1 y2 y3 = x1 x2 x3 x4 x5, testtwo het b(o) reps(999) a(0.05) seffi
> cient(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 70 data points relative to
 the bootstrapped reference set

 Smoothed heterogeneous bootstrap (999 replications)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50 (dots omitted)
```

CCR81: Nonparametric test of returns to scale (cont.)

```
. 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

Data: PWT5.6

- Penn World Tables
- Output: real GDP
- Inputs: capital stock and labour force

Malmquist productivity index decomposition

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

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

Improvements in productivity from period b to period c occur if $MPI_k^{o,bc} > 1$ ($MPI_k^{i,bc} < 1$)

Malmquist productivity index decomposition (cont.)

MPI maybe decomposed as

$$
\begin{aligned}
MPI_k^{o,bc} &= \frac{F^o(y_{k,b}, x_{k,b}, y_b, x_b|\text{CRS})}{F^o(y_{k,c}, x_{k,c}, y_c, x_c|\text{CRS})} \\
&\times \left[\frac{F^o(y_{k,c}, x_{k,c}, y_c, x_c|\text{CRS})}{F^o(y_{k,c}, x_{k,c}, y_b, x_b|\text{CRS})} \frac{F^o(y_{k,b}, x_{k,b}, y_c, x_c|\text{CRS})}{F^o(y_{k,b}, x_{k,b}, y_c, x_c|\text{CRS})} \right]^{1/2} \\
&= EFF \times TECH
\end{aligned}
$$

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).

If $EFF > 1$ (< 1 in input-based measurement), change in efficiency has positively contributed to 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.

Malmquist productivity index decomposition (cont.)

Calculating the Debreu-Farrell measure under VRS, MPI can be decomposed into three components attributable to (i) Pure Technical Efficiency Change (*PEFF*), (ii) Technological Change (*TECH*) and (iii) Scale Efficiency Change (*SEC*):

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

where S_k^o is scale efficiency.

Malmquist productivity index decomposition (cont.)

- It is not clear which of the decompositions should be used.
- We first perform the nonparametric test of returns to scale using heterogeneous bootstrap:

Malmquist productivity index decomposition (cont.)

```
. nptestrts y1965 = k1965 l1965, het b(o) reps(999) a(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) = 0.9920:
 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
 ------------------------------------------------------------------------------
```

Malmquist productivity index decomposition (cont.)

- We use the first decomposition

```
. teradial y1965 = k1965 l1965 (y1965 = k1965 l1965), r(c) b(o) te(F11) noprint
. teradial y1990 = k1990 l1990 (y1965 = k1965 l1965), r(c) b(o) te(F21) noprint
. teradial y1965 = k1965 l1965 (y1990 = k1990 l1990), r(c) b(o) te(F12) noprint
. teradial y1990 = k1990 l1990 (y1990 = k1990 l1990), r(c) b(o) te(F22) noprint
. g mpi = sqrt(F12 / F22 * F22 / F21)
. g effch = F11 / F22
. g techch = mpi / effch
```

- Reference specification in
  `y1965 = k1965 l1965 (y1965 = k1965 l1965)`
  and
  `y1990 = k1990 l1990 (y1990 = k1990 l1990)`
  is actually obsolete

Malmquist productivity index decomposition (cont.)

Table: Measures of technical efficiency and Malmquist Productivity Index for selected countries

#	Country	1965[a]	1990[b]	MPI	EFFch	TECHch
1	Argentina	1.000	1.546	0.818	0.647	1.264
2	Australia	1.320	1.213	1.184	1.088	1.088
3	Austria	1.174	1.374	1.067	0.854	1.249
4	Belgium	1.419	1.159	1.247	1.225	1.018
5	Bolivia	2.002	2.457	0.948	0.815	1.163
11	Equador	2.664	2.756	0.961	0.966	0.994
18	Hong Kong	2.202	1.000	1.519	2.202	0.690
32	New Zealand	1.186	1.406	1.004	0.843	1.191
33	Norway	1.628	1.257	1.492	1.295	1.152
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.

Outline

1. Nonparametric frontier analysis
   - Motivation
   - Radial efficiency analysis
   - Nonradial efficiency analysis
2. Statistical inference in radial frontier model
   - Type of the bootstrap for statistical inference
   - Returns to scale and scale analysis
3. Stata commands
   - `tenonradial`, `teradial`, `teradialbc`, `nptestind`, and `nptestrts`
4. Empirical application
   - Data: CCR81
   - Data: PWT5.6
5. Discussion
   - Sample restriction and runtime
   - Comparison to dea command in Stata, Stata Journal, 10(2): 267-80
   - Concluding remarks

Technical notes

Technical note

All functions create Stata matrices and feed them to plugin. The number of data points that can be used in all functions is thus limited by [R] **matsize**. Stata/IC allows maximum of 800, while Stata/MP and Stata/SE allow 11000 data points.

Technical note

Stata 11.2 and above can be used to run all new commands. Earlier versions of Stata can probably also be used, but Stata 11.2 is the earliest version available to authors.

Technical note

Solving linear programming problems make use of *quickhull* (`http://www.qhull.org/`, Barber et al. (1996)) algorithm 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 from C code. Systems for which the plugins are available are MacOX, Ubuntu, and Windows.

Runtime

```
. timer clear

. timer on 1

. use ccr81, clear
(Program Follow Through at 70 US Primary Schools)

. gen dref = x5 != 10

. tenonradial y1 y2 y3 = x1 x2 x3 x4 x5, r(c) b(o) ref(dref) tename(TErdCRSo) n
> oprint

. timer off 1

. timer on 2

. nptestind y1 y2 y3 = x1 x2 x3 x4 x5, r(c) b(o) reps(999) a(0.05) noprint

. timer off 2

. timer on 3

. teradialbc y1 y2 y3 = x1 x2 x3 x4 x5, r(v) b(o) ref(dref) reps(999) tebc(TErd
> VRSoBC1) biassqvar(TErdVRSoBC1bv) telower(TErdVRSoLB1) teupper(TErdVRSoUB1) n
> oprint

. timer off 3

. timer on 4

. teradialbc y1 y2 y3 = x1 x2 x3 x4 x5, het r(v) b(o) ref(dref) reps(999) tebc(
> TErdVRSoBC2) biassqvar(TErdVRSoBC2bv) telower(TErdVRSoLB2) teupper(TErdVRSoUB
> 2) noprint

. timer off 4

. timer on 5

. nptestrts y1 y2 y3 = x1 x2 x3 x4 x5, b(o) reps(999) a(0.05) sefficient(SEffnt
> _hom) noprint

. nptestrts y1 y2 y3 = x1 x2 x3 x4 x5, het b(o) reps(999) a(0.05) sefficient(SE
> ffnt_het) noprint

. timer off 6

. timer list
   1:      0.16 /        1 =       0.1600
   2:      2.80 /        1 =       2.8040
   3:      9.37 /        1 =       9.3670
   4:     20.23 /        1 =      20.2280
   5:     68.55 /        1 =      68.5490
   6:    782.47 /        1 =     782.4680
```

Comparison to dea command in Stata, Stata Journal, 10(2): 267-80

- Only radial measure
- Cannot be computed relative to a frontier formed by data points other than those for which measures are computed
- Slow

Comparison to dea command in Stata, Stata Journal, 10(2): 267-80 (cont.)

```
. 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´, r(v) b(i) te(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
```

Outlook

- Stata Journal: "*Nonparametric Frontier Analysis using Stata*"
- Stata command `simarwilson` (presented at German Stata Meeting 2015) can use it
  - For the first, but also for the second algorithm
- Statistical inference for nonradial measure

The End