第六届Stata中国用户大会·主题分享

绿色全要素生产率与高质量发展评估

张宁（山东大学）
zn928@naver.com

内容提要

- 背景
- 基本概念
- 模型
  - 数据包络分析(DEA)
  - 参数线性规划(PLP)
  - 随机前沿分析(SFA)
- Stata命令与示例
- 延伸

1. 背景：高质量发展内涵

- 2017年党的第十九次全国代表大会首次提出的新表述，表明中国经济，表明中国经济由高速增长阶段转向高质量发展阶段。
- 包括但不局限于:
  - 经济维度
  - 社会维度
  - 环境维度
  - ...
- 如何度量我国的经济建设成效?

1. 背景：全要素生产率的现实意义

- 2015年《政府工作报告》
  - 提高**全要素生产率**，加强质量、标准和品牌建设，加快培育新的增长点和增长极。
- 十九大报告
  - 坚持质量第一、效益优先，以供给侧结构性改革为主线，推动经济发展质量变革、效率变革、动力变革，提高**全要素生产率**。
- 2021年政府工作报告
  - 全员**劳动生产率**增长高于国内生产总值增长。

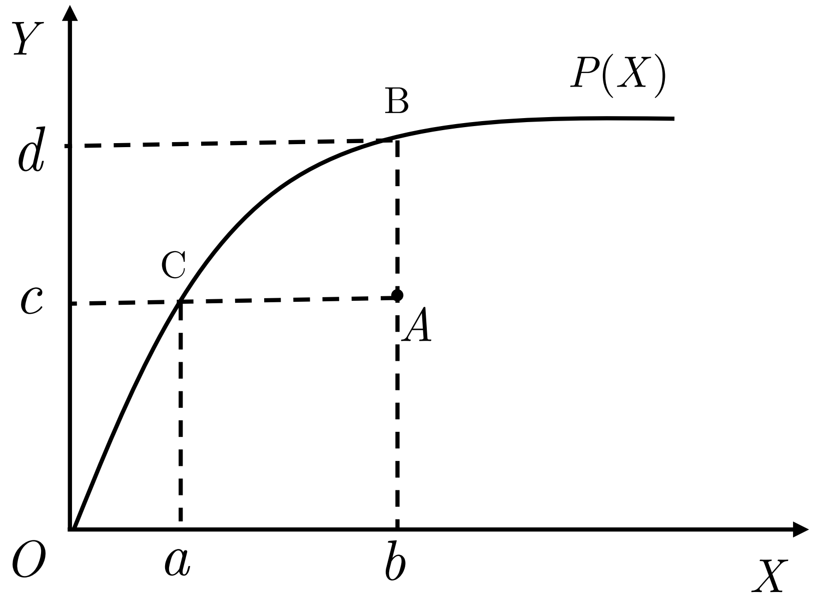
1. 背景：绿色全要素生产率

- 环境问题尖锐，传统TFP已不在适用于当前的需要
- 绿色全要素生产率(Green TFP)是衡量一个国家或区域资源、环境与经济协调发展的重要指标。
- 绿色TFP符合高质量发展的时代需求。
- 如何将环境变量纳入TFP测算中？是作为投入？还是作为产出？
  - 投入：$y=\beta X+\beta_b b+\alpha$。$X$是传统投入向量，$b$是环境变量。副产品是投入?
  - 产出：$(y,b)=\beta X+\alpha$。多产出的生产函数如何计算?
- Chung et al. (1997)提出了一种科学的度量方法——方向距离函数（DDF）

Productivity and Undesirable Outputs: A Directional Distance Function Approach
Authors: Y.H. Chung, R. Färe, S. Grosskopf
Source: *Journal of Environmental Management* 1997, 51: 229-240.

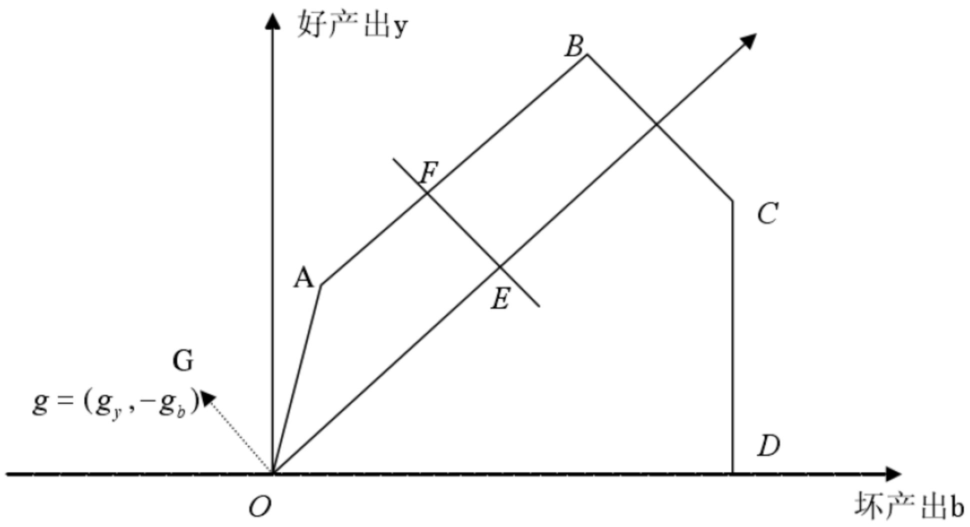
2. 基本概念

- 决策单元(DMU)：A点、B点
  - 企业、城市、部门、省份、国家等
- 生产前沿(Production frontier)：$P(X)$
- 生产可能性集合(PPS)/环境生产技术
- $T=\{(x,y,b): x\ can\ produce\ (y,b)\}$
  - x,y,b分别代表投入、期望产出以及非期望产出



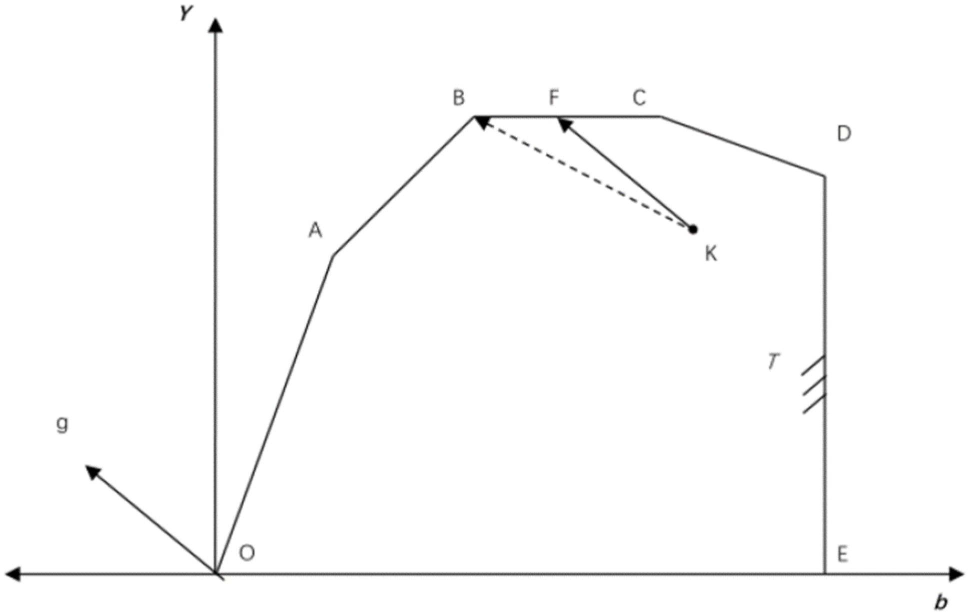
2. 基本概念

- **方向距离函数(DDF)是用于效率评估以及影子价格测算的一种最为常见的能源与环境建模技术，它可以同时为期望产出与非期望产出进行建模。**
- $\overrightarrow{D}(x, y, b; g) = sup\{\beta: ((x, y, b) + \beta g) \in T\}$
  - g**为方向向量，例如**$g = (-x, y, -b)$**或**$g = (-1, 1, -1)$



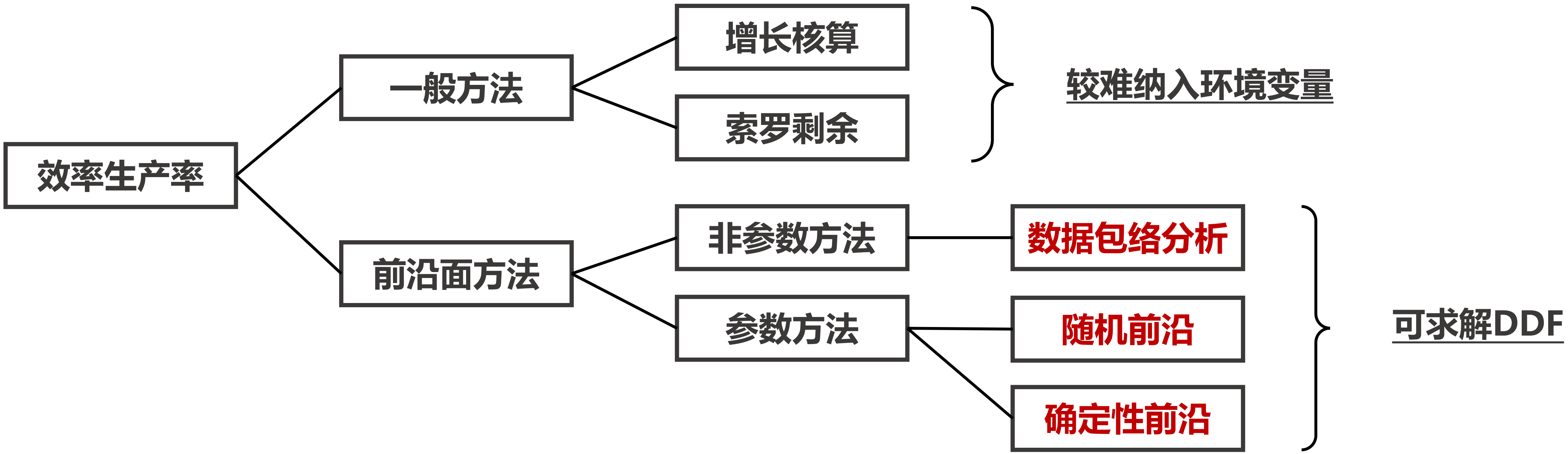
2. 基本概念

- 非径向方向距离函数(NDDF)
  - 相较于DDF：要素的非比例变化；松弛变量；可以度量要素(如能源)效率
- $\vec{D}(x, y, b; g) = sup\{w^T\beta: ((x, y, b) + diag(\beta)g) \in T\}$



3. 模型

- 如何对效率与生产率进行估算?



3. 模型

参数方法与非参数方法的优劣

- 数据包络分析DEA
  - 优点：无需价格信息、适用于SDF和DDF
  - 缺点：不可微分、易受极端值影响、无统计意义
- 参数线性规划PLP
  - 优点：处处可微、适用于SDF和DDF
  - 缺点：无统计意义
- 随机前沿分析SFA
  - 优点：处处可微、考虑随机因素、有统计意义
  - 缺点：不适用于SDF、可能违背经济学假设，如单调性

方法	优点	缺点
参数方法	(1)有经济学理论和内涵； (2)多阶可导； (3)提供参数估计	假设多：函数形式假设、分布假设等
非参数方法	(1)无须假设； (2)计算简单； (3)应用性广	(1)缺乏经济学理论； (2)无法提供参数估计

4. Stata命令与示例——DEA

- ddfeff 用以估计包含非期望产出的DDF
- nddfeff 用以估计包含非期望产出的NDDF
  - 前者是径向Debreu-Farrell测度，后者是非径向Russell测度
  - 可以应用于不同的生产技术，例如序列、全域
- gtfpch 用以估计绿色TFP变化
  - Malmquist-Luenberger生产率指数
  - Luenberger生产率指数

4. Stata命令与示例——DEA

- ddfeff与nddfeff命令安装

```
net install ddfeff, from("https://raw.githubusercontent.com/kerrydu/ddfeff/master/")
```

```
copy https://codeload.github.com/kerrydu/ddfeff/zip/master ddfeff-master.zip
unzipfile ddfeff-master.zip
net install ddfeff, from(`c(pwd)'/ddfeff-master)
```

- gtfpch命令安装

```
** install from github
net install gtfpch, from("https://raw.githubusercontent.com/kerrydu/gtfpch/master/") replace

**install from gitee
net install gtfpch, from(https://gitee.com/kerrydu/gtfpch/raw/master/) replace
```

4. Stata命令与示例——DEA

■ ddfeff命令

ddfeff *inputvars* = *desirable_outputvars* : *undesirable_outputvars* [if] [in], dmu(*varname*)

[time(*varname*) gx(*varlist*) gy(*varlist*) gb(*varlist*) sequential global vrs maxiter(#)

saving(*filename* [, *replace*])]

Title

ddfeff — Directional Distance Function for Efficiency/Productivity Analysis in Stata

Syntax

ddfeff *inputvars* = *desirable_outputvars* : *undesirable_outputvars* [*if*] [*in*] , **dmu**(*varname*) [*options*]

options	Description
Main	
dmu(**varname**)	specifies names of DMUs. It is required.
time(**varname**)	specifies time period for contemporaneous production technology. If **time**(*varname*) is not specified, global production technology is assumed.
gx(*varlist*)	specifies direction components for input adjustment. The default is gx=0.
gy(*varlist*)	specifies direction components for desirable output adjustment. The default is gy=Y.
gb(*varlist*)	specifies direction components for undesirable output adjustment. The default is gb=-B.
sequential	specifies sequential production technology.
glob	specifies global production technology.
productivity	specifies computing Malmquist-Luenberger productivity index.
vrs	specifies production technology with variable returns to scale. By default, production technology with constant returns to scale is assumed.
saving(*filename*[,*replace*])	specifies that the results be saved in *filename*.dta.

4. Stata命令与示例——DEA

- **nddfeff命令**

nddfeff *inputvars* = *desirable_outputvars* : *undesirable_outputvars* [if] [in], dmu(*varname*)

[time(*varname*) wmat(*name*) gx(*varlist*) gy(*varlist*) gb(varlist) sequential vrs

maxiter(#) saving(*filename* [, *replace*])]

Title

nddfeff — Non-radial Directional Distance Function in Stata

Syntax

nddfeff *inputvars* **=** *desirable_outputvars* **:** *undesirable_outputvars* [*if*] [*in*] **,** **dmu**(*varname*) [*options*]

options	Description
Main	
dmu(varname)	specifies names of DMUs. It is required.
time(varname)	specifies time period for contemporaneous production technology. If **time**(*varname*) is not specified, global production technology is assumed.
wmat(*name*)	specifies a weight matrix. The default is W=(1/(3*#.inputvars),..,1/(3*#.desirab > le_outputvars),...,1/(3*#.undesirabl > e_outputvars)).
gx(*varlist*)	specifies direction components for input adjustment. The default is gx=0.
gy(*varlist*)	specifies direction components for desirable output adjustment. The default is gy=Y.
gb(*varlist*)	specifies direction components for undesirable output adjustment. The default is gb=-B.
sequential	specifies sequential production technology.
vrs	specifies production technology with variable returns to scale. By default, production technology with constant returns to scale is assumed.
saving(*filename*[,*replace*])	specifies that the results be saved in *filename*.dta.

4. Stata命令与示例——DEA

- gtfpch命令

gtfpch *inputvars* = *desirable_outputvars* : *undesirable_outputvars* [if] [in],

[dmu(*varname*) luenberger ort(*string*) wmat(*name*) gx(*varlist*) gy(*varlist*) gb(*varlist*)

nonradial window(#) biennial sequential vrs global fgnz rd tol(*real*) maxiter(#)

saving(*filename* [, *replace*])]

Title

gtfpch — Total Factor Productivity with Undesirable Outputs in Stata

Syntax

gtfpch *inputvars* = *desirable_outputvars* : *undesirable_outputvars* [*if*] [*in*] , [*options*]

options	Description
Main	
dmu(varname)	specifies names of DMUs.
gx(*varlist*)	specifies direction components for input adjustment. The order of variables specified in gx() should as the same in *inputvars*.
gy(*varlist*)	specifies direction components for desirable output adjustment. The order of variables specified in gy() should as the same in *desirable_outputvars*.
gb(*varlist*)	specifies direction components for undesirable output adjustment. The order of variables specified in gb() should as the same in *undesirable_outputvars*.
sequential	specifies sequential production technology.
biennial	specifies biennial production technology.
global	specifies global production technology.
nonradial	specifies using non-radial directional distance function(NDDF).
*__wmat(*name*)__	specifies a weight matrix for adjustment of input and output variables.
***luenberger**	specifies estimating Luenberger productivity index. The default is Malmquist-Luenberger productivity

4. Stata命令与示例——DEA

- ## example_ddf.dta数据集

```
. use example_ddf, clear

.
end of do-file

. describe

Contains data from example_ddf.dta
  obs:            60
 vars:             7                          31 Oct 2019 20:31
-----------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
-----------------------------------------------------------------
gdp             double  %10.0g                gdp1997
co2             double  %10.0g                co2
labor           double  %10.0g                labor
capital         double  %10.0g                capital
energy          double  %10.0g                energy
id              byte    %10.0g                id1
t               byte    %10.0g                t
-----------------------------------------------------------------
Sorted by: id  t
```

4. Stata命令与示例——DEA

▪ DDF估计

```
. use example_ddf, clear

. ddfeff labor capital energy= gdp: co2, dmu(id) sav(ddf_result, replace)


 Directional Distance Function Results:
    (_Row: Row # in the original data; Dval: Estimated value of DDF)
```

	_Row	id	Dval
1.	1	1	-3.760e-13
2.	2	1	-4.115e-11
3.	3	2	.03664476
4.	4	2	.00119962
5.	5	3	.08231728
6.	6	3	.03871435
7.	7	4	.30107272
8.	8	4	.33846146
9.	9	5	-2.262e-09
10.	10	5	-1.592e-10
11.	11	6	.14955061
12.	12	6	.20002129
13.	13	7	.70741037

$$\vec{D}(x, y, b; g) = max\beta$$

$$s.t.\ \sum_{n=1}^{N} z_n x_n \le x - \beta g_x;$$

$$\sum_{n=1}^{N} z_n y_n \ge y + \beta g_y;$$

$$\sum_{n=1}^{N} z_n b_n = b - \beta g_b;$$

$$z_n \ge 0;\ \beta \ge 0$$

4. Stata命令与示例——DEA

▪ NDDF估计(Zhou et al., 2012; Zhang and Choi, 2013)

```
. nddfeff labor capital energy= gdp: co2, dmu(id) time(t) sav(nddf_result, replace)


 Non-raidal Directional Distance Function (NDDF) Results:
    (_Row: Row # in the original data; Dval: Estimated value of NDDF.)
```

	_Row	id	t	Dval	B_labor	B_capital	B_energy	B_gdp	B_co2
1.	1	1	9	.24623312	.61574767	.10973503	.21559886	7.652e-16	.42500551
2.	2	1	10	.24840085	.62711309	.1311299	.21144199	8.655e-15	.42197421
3.	3	2	9	6.424e-13	2.096e-16	4.023e-12	1.758e-12	8.834e-17	4.377e-19
4.	4	2	10	.05095057	4.165e-12	.32490274	.12563158	2.434e-12	.0026736
5.	5	3	9	.08516267	.18000491	.18361217	.04919117	9.129e-13	.11788526
6.	6	3	10	.08122075	.15788438	.20410304	.0469295	2.101e-10	.10735661
7.	7	4	9	.42472427	.59723414	1.517e-13	.5263787	.31573883	.58389638
8.	8	4	10	.43087165	.57959146	1.364e-13	.51791025	.34368102	.58310004
9.	9	5	9	5.398e-12	6.321e-12	1.745e-12	6.573e-12	1.047e-14	1.130e-11
10.	10	5	10	3.412e-12	4.688e-12	1.120e-12	4.170e-12	8.405e-15	6.903e-12
11.	11	6	9	.22544672	.65769858	.13135974	.27566896	8.993e-13	.32143106
12.	12	6	10	.23336103	.64372413	.17787968	.27855671	3.773e-13	.33336291
13.	13	7	9	.61815605	.68185012	7.225e-13	.52372976	.76777025	.68483793
14.	14	7	10	.62659101	.67180486	1.682e-12	.52204271	.77271388	.70910995
15.	15	8	9	.11982225	.45674362	.52874584	.04880903	6.786e-10	.01470059
16.	16	8	10	.14986929	.45731221	.54003207	.06634396	2.284e-13	.09504513

$$\overrightarrow{ND}(x, y, b; g) = \max\left(w_x\beta_x + w_y\beta_y + w_b\beta_b\right)$$

$$s.t.\quad \sum_{n=1}^{N} z_n x_n \le x - \beta_x g_x;$$

$$\sum_{n=1}^{N} z_n y_n \ge y + \beta_y g_y;$$

$$\sum_{n=1}^{N} z_n b_n = b - \beta_b g_b;$$

$$z_n \ge 0;\ \beta_x, \beta_y, \beta_b \ge 0$$

Energy and CO2 Emission Performance in Electricity Generation: A Non-radial Directional Distance Function Approach
Author: P. Zhou, B. W. Ang, H. Wang
Source: *European Journal of Operational Research* 2012, 221: 625-635.

Total-factor Carbon Emission Performance of Fossil Fuel Power Plants in China: A Metafrontier Non-radial Malmquist Index Analysis
作者：Ning Zhang, Yongrok Choi
来源：*Energy Economics* 2013, 40: 549-559.

4. Stata命令与示例——DEA

▪ Malmquist-Luenberger生产率指数

```
. gtfpch labor capital energy= gdp: co2, sav(ddf_result, replace)
New version available, 4.1 =>4.2
It can be updated by:
  net install gtfpch,from(https://raw.githubusercontent.com/kerrydu/gtfpch/master/) replace
or,
  net install gtfpch,from(https://gitee.com/kerrydu/gtfpch/raw/master/) replace

 The directional vector is (0 0 0 gdp -co2)


 Total Factor Productivity Change:Malmquist-Luenberger Productivity Index
    (Row: Row # in the original data; Pdwise: periodwise)
```

	Row	id	Pdwise	TFPCH	TECH	TECCH
1.	2	1	9~10	.	1.0000	.
2.	4	2	9~10	1.0359	0.9988	1.0371
3.	6	3	9~10	1.0435	0.9937	1.0501
4.	8	4	9~10	0.9714	0.9822	0.9890
5.	10	5	9~10	1.0207	1.0000	1.0207
6.	12	6	9~10	0.9625	0.9579	1.0047
7.	14	7	9~10	0.9937	0.9880	1.0058
8.	16	8	9~10	0.9782	0.9595	1.0194
9.	18	9	9~10	0.9733	0.9866	0.9865
10.	20	10	9~10	0.9280	0.9364	0.9911

$$M_t^{t+1}=\left[\frac{1+\vec{D}^t(x^t,y^t,b^t)}{1+\vec{D}^t(x^{t+1},y^{t+1},b^{t+1})}\times\frac{1+\vec{D}^{t+1}(x^t,y^t,b^t)}{1+\vec{D}^{t+1}(x^{t+1},y^{t+1},b^{t+1})}\right]^{0.5}$$

$$M_t^{t+1}=\frac{1+\vec{D}^t(x^t,y^t,b^t)}{1+\vec{D}^{t+1}(x^{t+1},y^{t+1},b^{t+1})}\times\left[\frac{1+\vec{D}^{t+1}(x^t,y^t,b^t)}{1+\vec{D}^t(x^t,y^t,b^t)}\times\frac{1+\vec{D}^{t+1}(x^{t+1},y^{t+1},b^{t+1})}{1+\vec{D}^t(x^{t+1},y^{t+1},b^{t+1})}\right]^{0.5}$$

4. Stata命令与示例——DEA

▪ Luenberger生产率指数

```
. gtfpch labor capital energy= gdp: co2, nonr luen sav(ddf_result, replace)

 The weight vector is (0 0 0 1 1)

 The directional vector is (0 0 0 gdp -co2)


 Total Factor Productivity Change:Luenberger Productivity Index (based on nonrial DDF)
    (Row: Row # in the original data; Pdwise: periodwise)
```

	Row	id	Pdwise	TFPCH	TECH	TECCH
1.	2	1	9~10	0.0155	-0.0083	0.0237
2.	4	2	9~10	.	-0.0027	.
3.	6	3	9~10	0.0518	0.0113	0.0404
4.	8	4	9~10	-0.0149	-0.0271	0.0123
5.	10	5	9~10	.	0.0000	.
6.	12	6	9~10	-0.0057	-0.0355	0.0297
7.	14	7	9~10	-0.0240	-0.0292	0.0052
8.	16	8	9~10	-0.0459	-0.0869	0.0410
9.	18	9	9~10	0.0006	-0.0118	0.0124
10.	20	10	9~10	-0.0406	-0.0584	0.0177
11.	22	11	9~10	0.0129	-0.0045	0.0174
12.	24	12	9~10	-0.0069	-0.0325	0.0256
13.	26	13	9~10	0.0314	0.0076	0.0238

$$LPI_{t,t+1} = \frac{1}{2}\left[\begin{array}{c}\left(\vec{D}_t(x^t,y^t,b^t;g) - \vec{D}_t(x^{t+1},y^{t+1},b^{t+1};g)\right)\\ +\left(\vec{D}_{t+1}(x^t,y^t,b^t;g) - \vec{D}_{t+1}(x^{t+1},y^{t+1},b^{t+1};g)\right)\end{array}\right]$$

$$E_{t,t+1} = \left(\vec{D}_t(x^t,y^t,b^t;g) - \vec{D}_{t+1}(x^{t+1},y^{t+1},b^{t+1};g)\right)$$

$$TC_{t,t+1} = \frac{1}{2}\left[\begin{array}{c}\left(\vec{D}_{t+1}(x^t,y^t,b^t;g) - \vec{D}_t(x^t,y^t,b^t;g)\right)\\ +\left(\vec{D}_{t+1}(x^{t+1},y^{t+1},b^{t+1};g) - \vec{D}_t(x^{t+1},y^{t+1},b^{t+1};g)\right)\end{array}\right]$$

4. Stata命令与示例——PLP

- dflp 用以使用线性规划技术估算距离函数
  - 投入距离函数
  - 产出距离函数
  - 方向距离函数
  - 异质性技术进步
  - 异质性生产技术
  - …

4. Stata命令与示例——PLP

- **dflp命令安装**

```
net install dflp, from("https://gitee.com/kerrydu/dflp/raw/master/") replace
```

4. Stata命令与示例——PLP

- **dflp命令**

dflp *inputvars* = *desirable_outputvars* : *undesirable_outputvars* [if] [in], dmu(*varname*)

[time(*varname*) input output directional normalize efficient maxiter(#)

saving(*filename* [, *replace*])]

Title

dflp — Estimating Input/Output/Directional Distance Function using linear programming techniques

Syntax

dflp *inputvars* = *desirable_outputvars* : *undesirable_outputvars* [*if*] [*in*] [,*options*]

options	Description
Main	
saving(*filename*[,*replace*])	specifies that the results be saved in *filename*.dta.
input	specifies estimating Input Distance Function.
output	specifies estimating Output Distance Function.
directional	specifies estimating Directional Distance Function. By default, DDF is assumed.
time(varname)	specifies the time period variable.
normalize	normalizes the variables.
efficient	computes the partial derivatives on the fronteir.
maxiter(#)	specifies the maximum number of iterations, which must be an integer greater than 0. The default value of maxiter is 16000.
tol(*real*)	specifies the convergence-criterion tolerance, which must be greater than 0. The default value of tol is 1e-8.

4. Stata命令与示例——PLP

- plp.dta数据集

```
. use plp.dta, clear

. describe

Contains data from plp.dta
  obs:            95
 vars:            49                          1 Sep 2021 16:06
------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
------------------------------------------------------------------------------
_Row            float   %9.0g                 Row # in the original data set
k               float   %8.0g                 K
l               float   %8.0g                 L
e               float   %8.0g                 E
y               float   %8.0g                 Y
b               float   %8.0g                 B
lnk             float   %9.0g                 log(k)
lnl             float   %9.0g                 log(l)
lne             float   %9.0g                 log(e)
lny             float   %9.0g                 log(y)
lnb             float   %9.0g                 log(b)
lnk_lnl         float   %9.0g                 lnk * lnl
lnk_lne         float   %9.0g                 lnk * lne
lnl_lne         float   %9.0g                 lnl * lne
```

4. Stata命令与示例——PLP

- **IDF估计**

```
.        dflp k l e = y: b, in
```

Parameters in Input Distance function:

Variable	Coefficient
_Cons	0.2254
lnk	0.5218
lnl	0.2051
lne	0.2732
lny	-0.7696
lnb	0.1884
lnk_lnl	0.1293
lnk_lne	0.6894
lnl_lne	0.2880
lnk2	-0.8188
lnl2	-0.2880
lne2	0.0000
lny2	-6.1630
lnb2	-0.6544
lny_lnb	-1.4750
lnk_lny	0.0737
lnl_lny	0.8516
lne_lny	-0.9253

超越对数

- $min \sum_{t=1}^{T}\sum_{k=1}^{K}\left[ln\left(D_i\left(x^{kt},y^{kt},b^{kt}\right)\right)-ln\,1\right]$
- $s.t.\,ln\left(D_i\left(x^{kt},y^{kt},b^{kt}\right)\right)\geq 0, k=1,2,\dots,K,$
- $ln\left(D_i\left(x^{kt},y^{kt},0\right)\right)<0, k=1,2,\dots,K,$
- $\frac{\partial ln\left(D_i\left(x^{kt},y^{kt},b^{kt}\right)\right)}{\partial ln\, x_m^{kt}}\geq 0, m=1,2,\dots,M,$
- $\frac{\partial ln\left(D_i\left(x^{kt},y^{kt},b^{kt}\right)\right)}{\partial ln\, y_n^{kt}}\leq 0, n=1,2,\dots,N,$
- $\frac{\partial ln\left(D_i\left(x^{kt},y^{kt},b^{kt}\right)\right)}{\partial ln\, b_j^{kt}}\geq 0, j=1,2,\dots,J,$
- $\sum_{m=1}^{M}\alpha_m^{x}=1;\ \sum_{m=1}^{M}\sum_{m1=1}^{M}\beta_{mm1}^{xx}=0;$
- $\sum_{m=1}^{M}\sum_{n=1}^{N}\beta_{mn}^{xy}=\sum_{m=1}^{M}\sum_{j=1}^{J}\beta_{mj}^{xb}=0;$
- $\beta_{mm1}^{xx}=\beta_{m1m}^{xx};\beta_{nn1}^{yy}=\beta_{n1n}^{yy};\beta_{jj1}^{bb}=\beta_{j1j}^{bb}.$

4. Stata命令与示例——PLP

ODF估计

```
.        dflp k l e = y: b, out
```

Parameters in Output Distance function:

Variable	Coefficient
_Cons	-0.0974
lnk	0.0219
lnl	-0.1312
lne	0.0187
lny	1.0568
lnb	-0.0568
lnk_lnl	0.2859
lnk_lne	0.2655
lnl_lne	-0.2892
lnk2	1.8902
lnl2	-0.1314
lne2	0.4149
lny2	-0.7482
lnb2	0.2068
lny_lnb	0.5415
lnk_lny	0.1504
lnl_lny	-0.0297
lne_lny	0.0496

超越对数

- $min \sum_{t=1}^{T}\sum_{k=1}^{K}\left[ln\left(D_o\left(x^{kt},y^{kt},b^{kt}\right)\right)-ln\,1\right]$
- $s.t.\, ln\left(D_o\left(x^{kt},y^{kt},b^{kt}\right)\right)\geq 0, k=1,2,\ldots,K,$
- $ln\left(D_o\left(x^{kt},y^{kt},0\right)\right)< 0, k=1,2,\ldots,K,$
- $\frac{\partial ln\left(D_o\left(x^{kt},y^{kt},b^{kt}\right)\right)}{\partial ln\, x_m^{kt}}\geq 0, m=1,2,\ldots,M,$
- $\frac{\partial ln\left(D_o\left(x^{kt},y^{kt},b^{kt}\right)\right)}{\partial ln\, y_n^{kt}}\leq 0, n=1,2,\ldots,N,$
- $\frac{\partial ln\left(D_o\left(x^{kt},y^{kt},b^{kt}\right)\right)}{\partial ln\, b_j^{kt}}\geq 0, j=1,2,\ldots,J,$
- $\sum_{n=1}^{N}\alpha_n^y+\sum_{j=1}^{J}\alpha_j^b=$ $1;\ \sum_{n=1}^{N}\sum_{n1=1}^{N}\beta_{nn1}^{yy}=\sum_{j=1}^{J}\sum_{j1=1}^{J}\beta_{jj1}^{bb}=0;$
- $\sum_{n=1}^{N}\sum_{j=1}^{J}\beta_{nj}^{yb}=\sum_{m=1}^{M}\sum_{n=1}^{N}\beta_{mn}^{xy}=\sum_{m=1}^{M}\sum_{j=1}^{J}\beta_{mj}^{xb}=0;$
- $\beta_{mm1}^{xx}=\beta_{m1m}^{xx};\beta_{nn1}^{yy}=\beta_{n1n}^{yy};\beta_{jj1}^{bb}=\beta_{j1j}^{bb}.$

4. Stata命令与示例——PLP

- ## DDF估计

```
.        dflp k l e = y: b, dir

Parameters in Directional Distance function:
```

Variable	Coefficient
_Cons	-0.5077
k	3.0287
l	-0.9674
e	0.5128
y	-0.9126
b	0.0874
k_l	-0.4629
k_e	-0.3855
l_e	0.2263
k2	-2.3824
l2	0.5869
e2	-0.4660
y2	-0.2095
b2	-0.2095
y_b	-0.2095
k_y	0.0824
l_y	0.3956
e_y	0.0300

二次型

- $min \sum_{t=1}^{T}\sum_{k=1}^{K}[\vec{D}(x^{kt}, y^{kt}, b^{kt}; g_y, -g_b) - 0]$
- $s.t. \vec{D}(x^{kt}, y^{kt}, b^{kt}; g_y, -g_b) \geq 0, k = 1, 2, \dots, K,$
- $\vec{D}(x^{kt}, y^{kt}, 0; g_y, -g_b) < 0, k = 1, 2, \dots, K,$
- $\frac{\partial\vec{D}(x^{kt}, y^{kt}, b^{kt}; g_y, -g_b)}{\partial \ln x_m^{kt}} \geq 0, m = 1, 2, \dots, M,$
- $\frac{\partial\vec{D}(x^{kt}, y^{kt}, b^{kt}; g_y, -g_b)}{\partial \ln y_n^{kt}} \leq 0, n = 1, 2, \dots, N,$
- $\frac{\partial\vec{D}(x^{kt}, y^{kt}, b^{kt}; g_y, -g_b)}{\partial \ln b_j^{kt}} \geq 0, j = 1, 2, \dots, J,$
- $g_y \sum_{n=1}^{N} \alpha_n^y - g_b \sum_{j=1}^{J} \alpha_j^b = -1;$
- $g_y \sum_{n=1}^{N} \sum_{n1=1}^{N} \beta_{nn1}^{yy} - g_b \sum_{n=1}^{N} \sum_{j=1}^{J} \beta_{nj}^{yb} = 0;$
- $g_y \sum_{n=1}^{N} \sum_{j=1}^{J} \beta_{nj}^{yb} - g_b \sum_{j=1}^{J} \sum_{j1=1}^{J} \beta_{jj1}^{bb} = 0;$
- $g_y \sum_{m=1}^{M} \sum_{n=1}^{N} \beta_{mn}^{xy} - g_b \sum_{m=1}^{M} \sum_{j=1}^{J} \beta_{jj}^{xb} = 0;$
- $\beta_{mm1}^{xx} = \beta_{m1m}^{xx}; \beta_{nn1}^{yy} = \beta_{n1n}^{yy}; \beta_{jj1}^{bb} = \beta_{j1j}^{bb}.$

4. Stata命令与示例——SFA

- frontier/xtfrontier/sfcross/sfpanel 用以利用SFA估算技术效率
  - 方向距离函数
  - 异质性生产技术
  - 能源效率
  - 影子价格
  - …

4. Stata命令与示例——SFA

- frontier/xtfrontier命令安装
  - 官方命令，无需安装
- sfcross/sfpanel命令安装

```
***install sfcross
ssc install sfcross

***install sfpanel
ssc install sfpanel
```

Authors: Federico Belotti, Silvio Daidone, Vincenze Atella, Giuseppe Ilardi

4. Stata命令与示例——SFA

▪ frontier与sfcross命令——SFA for cross-sectional data

frontier *depvar* [*indepvars*] [if] [in] [*weight*], noconstant distribution(exponential)

distribution(tnormal) distribution(hnormal) ufrom(matrix) cm(*varlist* [, noconstant])

sfcross *depvar* [*indepvars*] [if] [in] [*weight*], noconstant distribution(exponential)

distribution(tnormal) distribution(hnormal) ufrom(matrix) cm(*varlist* [, noconstant])

frontier *depvar* [*indepvars*] [*if*] [*in*] [*weight*] [, *options*]

options	Description
Model	
noconstant	suppress constant term
distribution(hnormal)	half-normal distribution for the inefficiency term
distribution(exponential)	exponential distribution for the inefficiency term
distribution(tnormal)	truncated-normal distribution for the inefficiency term
ufrom(*matrix***)**	specify untransformed log likelihood; only with **d(tnormal)**
cm(*varlist***[, noconstant])**	fit conditional mean model; only with **d(tnormal)**; use **noconstant** to suppress constant term

sfcross *depvar* [*indepvars*] [*if*] [*in*] [*weight*] [, *options*]

options	Description
Frontier	
noconstant	suppress constant term
distribution(exponential)	exponential distribution for the inefficiency term, the default
distribution(hnormal)	half-normal distribution for the inefficiency term
distribution(tnormal)	truncated-normal distribution for the inefficiency term
distribution(gamma)	gamma distribution for the inefficiency term

4. Stata命令与示例——SFA

- **xtfrontier命令——SFA for panel data**

xtfrontier *depvar* [*indepvars*] [if] [in] [*weight*], ti [*ti_options*] **Time-invariant model**

xtfrontier *depvar* [*indepvars*] [if] [in] [*weight*], tvd [*tvd_options*] **Time-varying decay model**

ti_options	Description
Model	
noconstant	suppress constant term
ti	use time-invariant model
cost	fit cost frontier model
constraints(*constraints*)	apply specified linear constraints
SE	
vce(*vcetype*)	*vcetype* may be **oim**, **bootstrap**, or **jackknife**
Reporting	
level(#)	set confidence level; default is **level(95)**
nocnsreport	do not display constraints
display_options	control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
Maximization	
maximize_options	control the maximization process; seldom used
collinear	keep collinear variables
coeflegend	display legend instead of statistics

tvd_options	Description
Model	
noconstant	suppress constant term
tvd	use time-varying decay model
cost	fit cost frontier model
constraints(*constraints*)	apply specified linear constraints
SE	
vce(*vcetype*)	*vcetype* may be **oim**, **bootstrap**, or **jackknife**
Reporting	
level(#)	set confidence level; default is **level(95)**
nocnsreport	do not display constraints
display_options	control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
Maximization	
maximize_options	control the maximization process; seldom used
collinear	keep collinear variables
coeflegend	display legend instead of statistics

4. Stata命令与示例——SFA

▪ sfpanel命令——SFA for panel data

sfpanel *depvar* [*indepvars*] [if] [in] [*weight*], model(tfe) [*tfe_options*] **True fixed-effects model**

tfe_options	Description
Inefficiency distribution	
distribution(exponential)	exponential distribution for the inefficiency term, the default
distribution(hnormal)	half-normal distribution for the inefficiency term
distribution(tnormal)	truncated-normal distribution for the inefficiency term
Ancillary equations	
emean(*varlist_m*[, **noconstant**])	fit conditional mean model; only with **d(tnormal)**; use **noconstant** to suppress constant term
usigma(*varlist_u*[, **noconstant**])	specify explanatory variables for the inefficiency variance function; use **noconstant** to suppress constant term
vsigma(*varlist_v*[, **noconstant**])	specify explanatory variables for the idiosyncratic error variance function; use **noconstant** to suppress constant term

4. Stata命令与示例——SFA

- plp.dta数据集

```
. use rawdata.dta,clear

. d

Contains data from rawdata.dta
  obs:           558
 vars:            15                          2 Dec 2019 09:57
-------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------
name            str51   %51s                  Plant Name
firm            str9    %9s                   Firm
year            int     %10.0g                Year
metaID          byte    %10.0g                meta ID
groupID         byte    %10.0g                GroupID
t               byte    %10.0g                t
group           byte    %10.0g                group
Y               double  %10.0g                Y.electricity. Billion kwh
K               double  %10.0g                x1.Total Asset Value billion Y
L               int     %10.0g                x3.L.persons
E               double  %10.0g                x2.energy.1000t
```

4. Stata命令与示例——SFA

■ SFA对传统生产函数的估计

```
.        frontier Y K L E, d(h) // d(h)代表half-normal distribution, 半正态分布

Iteration 0:   log likelihood = -144.32454
Iteration 1:   log likelihood = -137.97658
Iteration 2:   log likelihood = -137.90501
Iteration 3:   log likelihood = -137.90498
Iteration 4:   log likelihood = -137.90498

Stoc. frontier normal/half-normal model         Number of obs     =        558
                                                Wald chi2(3)      =   46144.62
Log likelihood = -137.90498                     Prob > chi2       =     0.0000
```

Y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
K	.2657428	.0270835	9.81	0.000	.2126602	.3188254
L	-.0001242	.0000148	-8.38	0.000	-.0001532	-.0000952
E	.0268162	.0004517	59.37	0.000	.025931	.0277015
_cons	.3313874	.0432109	7.67	0.000	.2466956	.4160793
/lnsig2v	-3.122804	.1746842	-17.88	0.000	-3.465178	-2.780429
/lnsig2u	-1.887349	.1739858	-10.85	0.000	-2.228355	-1.546343
sigma_v	.2098417	.018328			.176826	.2490219
sigma_u	.3891951	.0338572			.3281851	.461547
sigma2	.1955064	.0211553			.1540427	.2369701
lambda	1.854708	.0491875			1.758302	1.951114

LR test of sigma_u=0: chibar2(01) = 18.63   Prob >= chibar2 = 0.000

- $Y = a_0 + a_L L + a_K K + a_E E + v - u$
- $TE = e^{-u}$

```
.        predict techeff if e(sample), te //TE=E[exp(-u)|e],

.

.        sum techeff, detail //基本统计量
```

technical efficiency via E[exp(-su)|e]

	Percentiles	Smallest		
1%	.3574572	.2618145		
5%	.5244772	.2756284		
10%	.602897	.318954	Obs	558
25%	.7080623	.3422965	Sum of Wgt.	558
50%	.7771713		Mean	.755848
		Largest	Std. Dev.	.1179138
75%	.8400869	.9380204		
90%	.8763755	.9424443	Variance	.0139037
95%	.899304	.9432959	Skewness	-1.310029
99%	.9359514	.9467551	Kurtosis	5.022244

4. Stata命令与示例——SFA

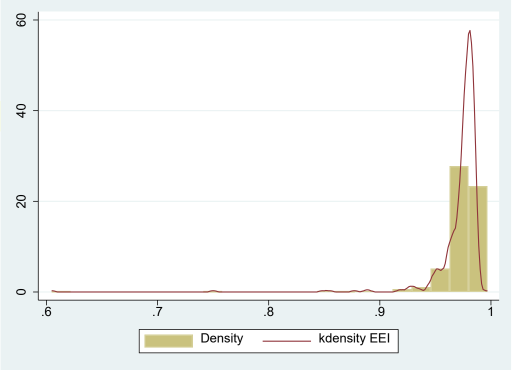
■ SFA对能源效率的估计

```
Stoc. frontier normal/tnormal model              Number of obs  =       558
                                                 Wald chi2(3)   =  27552.09
                                                 Prob > chi2    =    0.0000

Log likelihood =   887.9930
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Frontier						
LnK	.2127521	.0219388	9.70	0.000	.169753	.2557513
LnL	-.0162296	.0024411	-6.65	0.000	-.0210141	-.0114452
LnY	-1.21902	.020974	-58.12	0.000	-1.260128	-1.177912
_cons	-3.200848	.0261223	-122.53	0.000	-3.252047	-3.149649
Mu						
_cons	-38.87182	24.54375	-1.58	0.113	-86.97668	9.233033
Usigma						
_cons	.0907962	.6423036	0.14	0.888	-1.168096	1.349688
Vsigma						
_cons	-6.409945	.1026521	-62.44	0.000	-6.61114	-6.208751
sigma_u	1.046444	.3360675	3.11	0.002	.5576365	1.963727
sigma_v	.04056	.0020818	19.48	0.000	.0366783	.0448525
lambda	25.7999	.3363204	76.71	0.000	25.14073	26.45908

H0: No inefficiency component: z = 35.422   Prob<=z = 1.000

- $D_E(L_i, K_i, E_i, Y_i) = E_i D_E(L_i, K_i, 1, Y_i)$
- $\ln D_E(L_i, K_i, E_i, Y_i) = \ln E_i + \beta_0 + \beta_K \ln K_i + \beta_L \ln L_i + \beta_Y \ln Y_i + v_i$
- $\ln(1/E_i) = \ln E_i + \beta_0 + \beta_K \ln K_i + \beta_L \ln L_i + \beta_Y \ln Y_i + v_i - u_i$




```
use rawdata.dta,clear

gen y = -log(E) // Eq.(6)移项得出
gen LnK = log(K)
gen LnL = log(L)
gen LnY = log(Y) //根据文章Eq.(8)定义

global xvar LnK LnL LnY

sfcross y $xvar, d(t) // u呈截尾正态分布

help sfcross_postestimation //  Postestimation tools for sfcross

predict EEI, bc //estimates of energy efficiency via E[exp(-u)|e]

sum EEI, detail //基本统计量

twoway (histogram EEI) (kdensity EEI) //直方图和核密度函数图
```

4. Stata命令与示例——SFA

▪ Meta-SFA测算资本效率

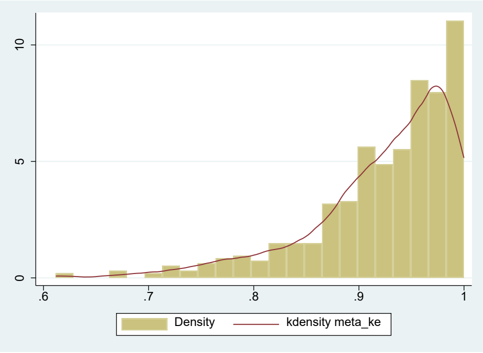
```
*** 国有企业 Group1 ***

    use rawdata.dta,clear

    keep  if group == 1 // 提取Group1
    xtset groupID year // 设置面板

    gen y = -lnK
    gen alphL = lnL
    gen alphE = lnE
    gen alphY = lnY
    gen betaLL = 0.5*lnL*lnL
    gen betaEE = 0.5*lnE*lnE
    gen betaYY = 0.5*lnY*lnY
    gen betatt = 0.5*t*t
    gen betaLE = lnL*lnE
    gen betaLY = lnL*lnY
    gen betaEY = lnE*lnY // translog函数形式求资本效率

    global xvar alphL alphE alphY betaLL betaEE betaYY betaLE betaLY betaEY

    sfpanel y $xvar, dis(h) m(tfe) difficult rescale // truncated-normal distr
fixed-effects model (Greene, 2005)

    help sfpanel_postestimation //  Postestimation tools for sfpanel

    estimates store Kgroup1

    predict groupke, bc // capital efficiency TE=E[exp(-u)|e] in Group1

    predict yhat, xb // 估计y的拟合值yhat

    sum groupke, detail //基本统计量

    save Group1.dta,replace
```

组内前沿

- $ln\, Y_{jit} = ln\, f_t^j(X_{jit}) + V_{jit} - U_{jit}$
- $ln\, \hat{f}_t^j(X_{jit}) = ln\, f_t^M(X_{jit}) - U_{jit}^M + V_{jit}^M$

共同前沿

```
*** Meta-SFA ***

    use Group.dta, replace
    xtset metaID year

    global xvar alphL alphE alphY betaLL betaEE betaYY betaLE betaLY betaEY

    sfpanel yhat $xvar, dis(h) m(tfe) difficult rescale //yhat估计  yhatgroup =
-u, u= lnTGD

    estimates store Kmeta

    predict tgr, bc   // 估计 TGR = 1/TGD = E[exp(-u)|e]

    gen meta_ke = groupke*tgr   // 计算Meta-frontier的资本效率

    sum groupke tgr meta_ke, detail //基本统计量

    twoway (histogram groupke) (kdensity groupke) //直方图和核密度函数图
    twoway (histogram tgr) (kdensity tgr) //直方图和核密度函数图
    twoway (histogram meta_ke) (kdensity meta_ke) //直方图和核密度函数图
```



5. 延伸

Measuring technical efficiency and total factor productivity change with undesirable outputs in Stata

Daoping Wang, Kerui Du, Ning Zhang

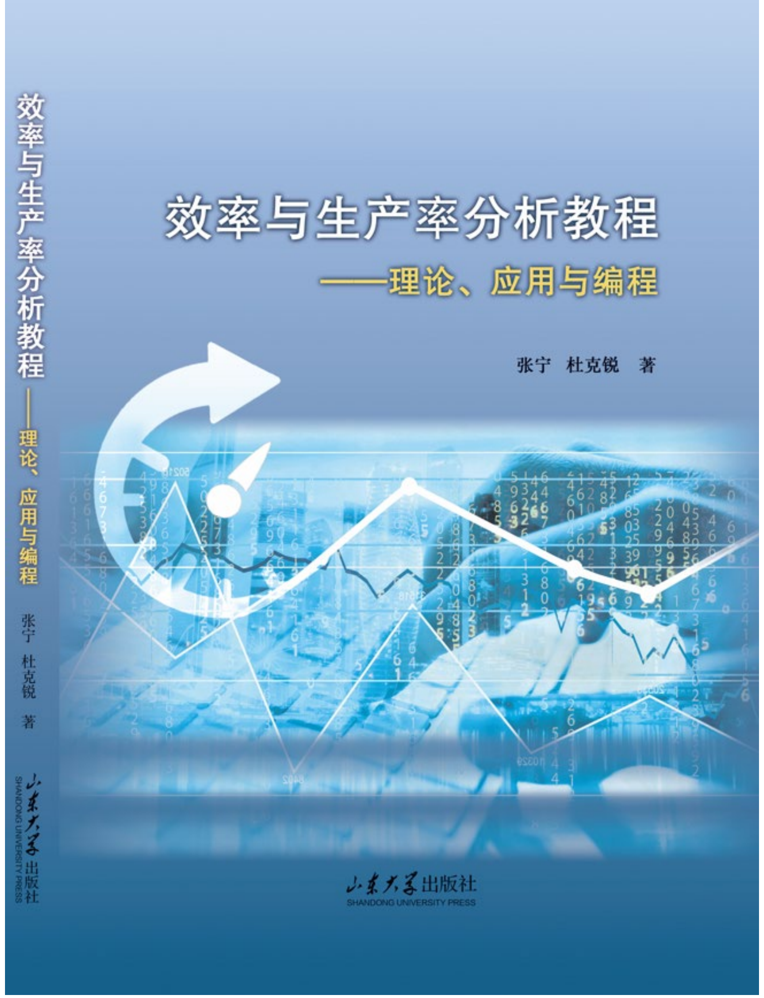
First Published April 5, 2022 | Research Article | Check for updates
https://doi.org/10.1177/1536867X221083886

The Stata Journal: Promoting communications on statistics and Stata

Article information

Abstract

In this article, we introduce two community-contributed data envelopment analysis commands for measuring technical efficiency and productivity change in Stata. Over the last decades, an important theoretical progression of data envelopment analysis, a nonparametric method widely used to assess the performance of decision-making units, is the incorporation of undesirable outputs. Models able to deal with undesirable outputs have been developed and applied in empirical studies for assessing the sustainability of decision-making units. These models are getting more and more attention from researchers and managers. The teddf command discussed in the present article allows users to measure technical efficiency, both radial and nonradial, when some outputs are undesirable. Technical efficiency measures are obtained by solving linear programming problems. The gtfpch command we also describe here provides tools for measuring productivity change, for example, the Malmquist–Luenberger index and the Luenberger indicator. We provide a brief overview of the nonparametric efficiency and productivity change measurement accounting for undesirable outputs, and we describe the syntax and options of the new commands. We also illustrate with examples how to perform the technical efficiency and productivity analysis with the newly introduced commands.



Thanks