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

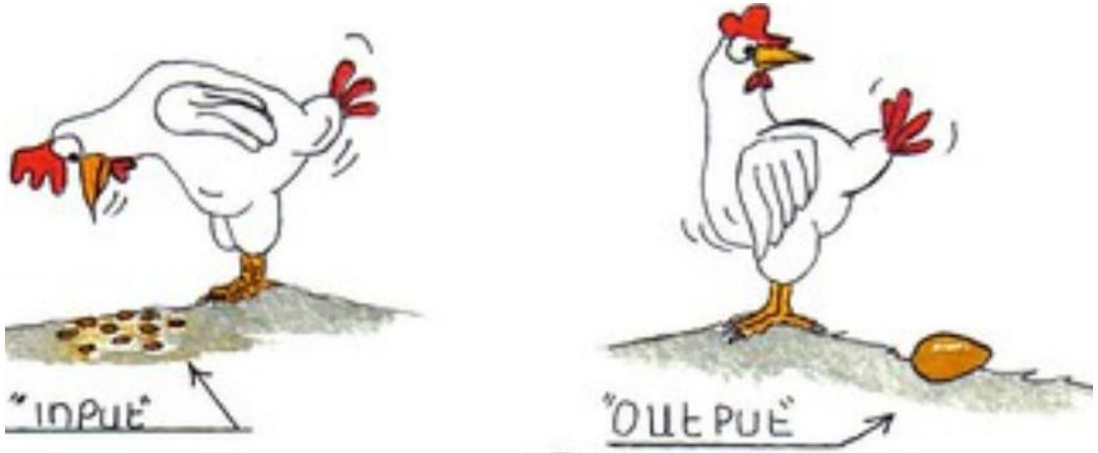
王道平
上海财经大学
DaopingWang@outlook.com

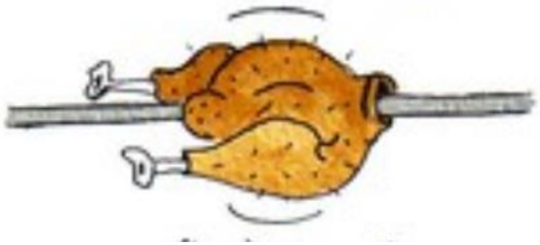
Outline

- Introduction
- Model
- Stata commands
- Illustrative example
- Outlook

Introduction

Efficiency



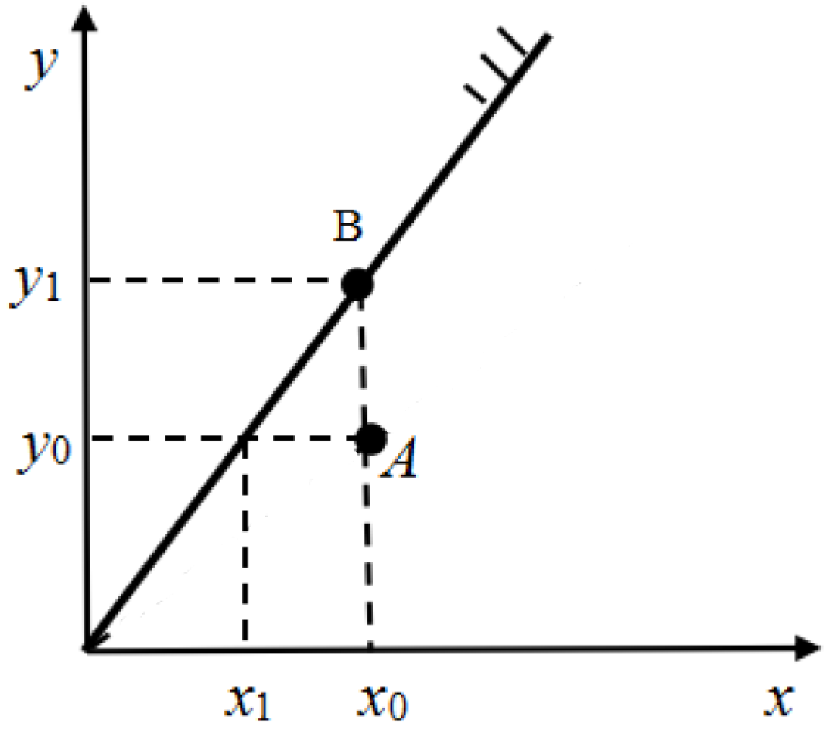


Introduction

Some basic concepts

- Decision-Making Unit (DMU)
  - firms, cities, provinces
- Production Frontier
- Production Possibility Set (PPS)/ Technology set

$$T = \{(\boldsymbol{x}, \boldsymbol{y}) : \boldsymbol{x} \text{ can produce } \boldsymbol{y}\}$$



Introduction

Distance function

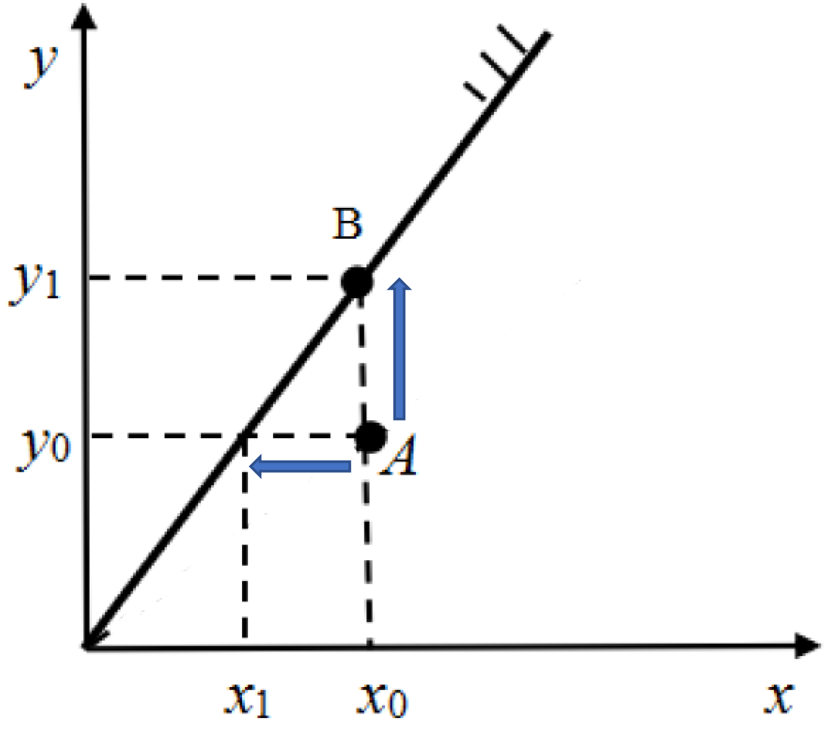
- The output-oriented measure

$$D_o(\boldsymbol{x},\boldsymbol{y}) = \sup\{\beta : ((\boldsymbol{x},\boldsymbol{y}) + \beta(\boldsymbol{0},\boldsymbol{y})) \in T\}$$

- The input-oriented measure

$$D_i(\boldsymbol{x},\boldsymbol{y}) = \sup\{\beta : ((\boldsymbol{x},\boldsymbol{y}) + \beta(-\boldsymbol{x},\boldsymbol{0})) \in T\}$$

$$D_r(\boldsymbol{x},\boldsymbol{y};\boldsymbol{g}) = \sup\{\beta : ((\boldsymbol{x},\boldsymbol{y}) + \beta\boldsymbol{g}) \in T\}$$

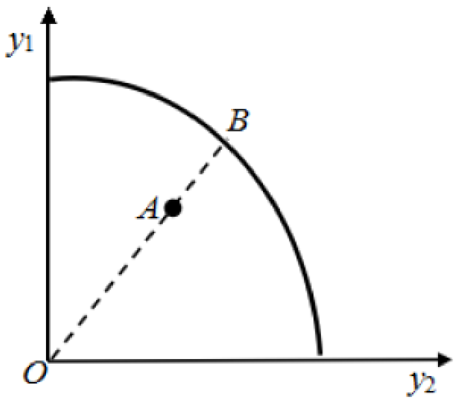


Introduction

Radial measure

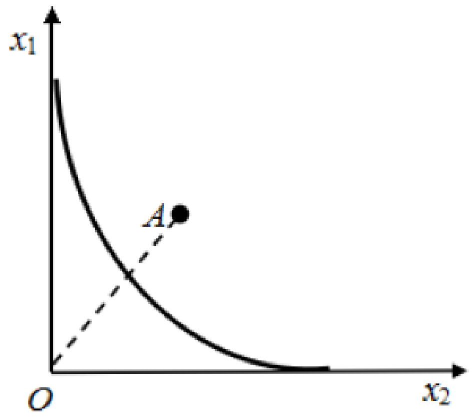
- The output-oriented measure

$$D_o(\boldsymbol{x}, \boldsymbol{y}) = \sup\{\beta : ((\boldsymbol{x}, \boldsymbol{y}) + \beta(\mathbf{0}, \boldsymbol{y})) \in T\}$$



- The input-oriented measure

$$D_i(\boldsymbol{x}, \boldsymbol{y}) = \sup\{\beta : ((\boldsymbol{x}, \boldsymbol{y}) + \beta(-\boldsymbol{x}, \mathbf{0})) \in T\}$$



Introduction

Slack in radial measure

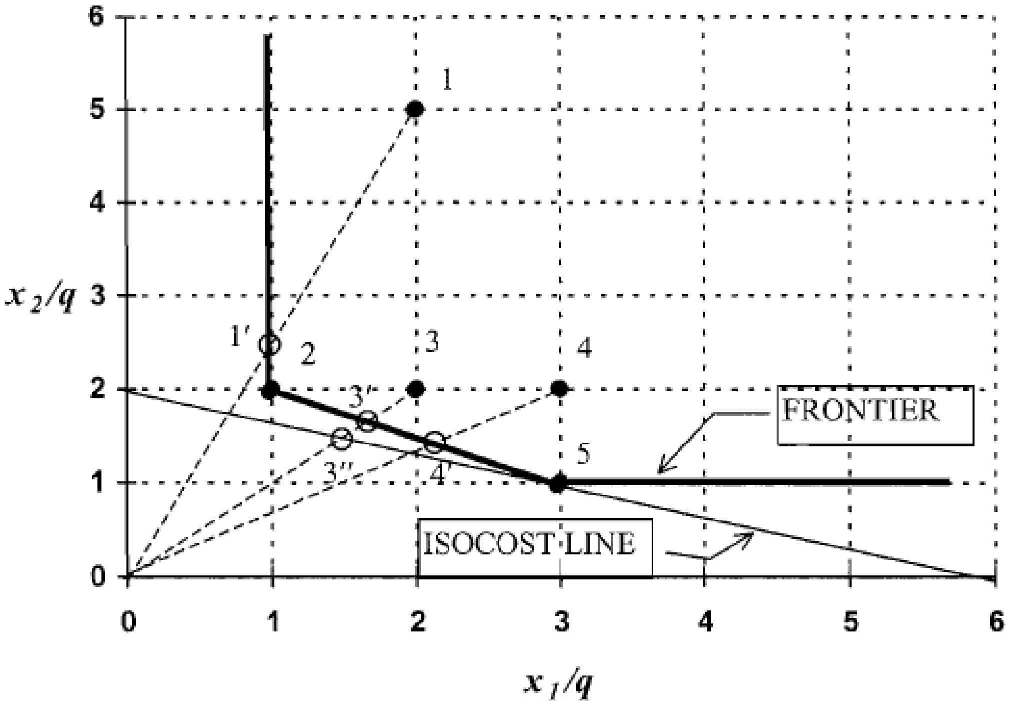
- The input-oriented measure

$$D_i(\boldsymbol{x},\boldsymbol{y}) = \sup\{\beta : ((\boldsymbol{x},\boldsymbol{y}) + \beta(-\boldsymbol{x},\boldsymbol{0})) \in T\}$$

Non-radial measure

$$D_{nr}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{g}) = \sup\{\boldsymbol{w}^T\boldsymbol{\beta} : ((\boldsymbol{x},\boldsymbol{y}) + diag(\boldsymbol{\beta})\cdot\boldsymbol{g}) \in T\}$$

$$D_r(\boldsymbol{x},\boldsymbol{y};\boldsymbol{g}) = \sup\{\beta : ((\boldsymbol{x},\boldsymbol{y}) + \beta\boldsymbol{g}) \in T\}$$



Introduction

Undesirable outputs

- Sustainability
- Environmental regulation

Introduction

Directional distance functions with undesirable outputs

- Technology set

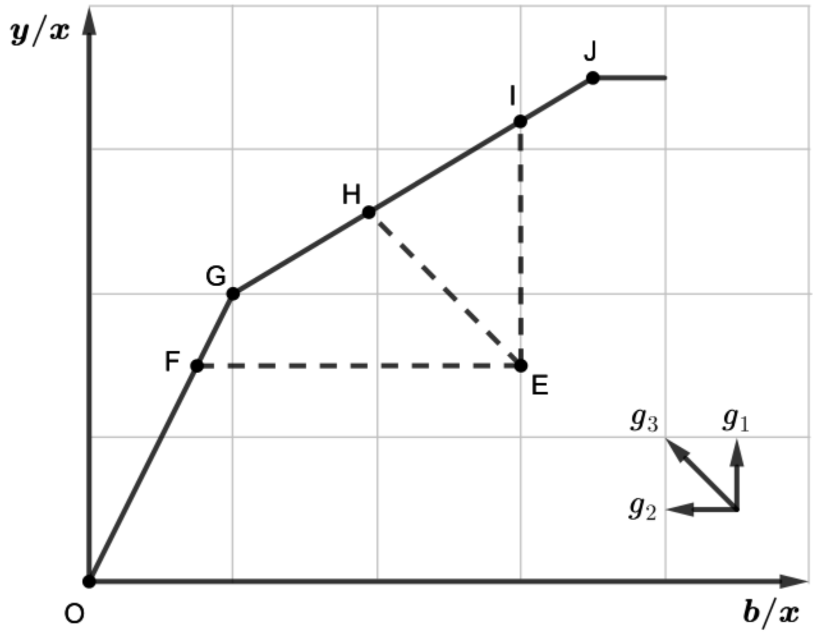
$$T = \{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{b}) : \boldsymbol{x} \text{ can produce } (\boldsymbol{y}, \boldsymbol{b})\}$$

- Radial measure

$$D_r(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{b}; \boldsymbol{g}) = \sup\{\beta : ((\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{b}) + \beta \boldsymbol{g}) \in T\}$$

- Non-radial measure

$$D_{nr}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{b}; \boldsymbol{g}) = \sup\{\boldsymbol{w}^T \boldsymbol{\beta} : ((\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{b}) + diag(\boldsymbol{\beta}) \cdot \boldsymbol{g}) \in T\}$$



Model

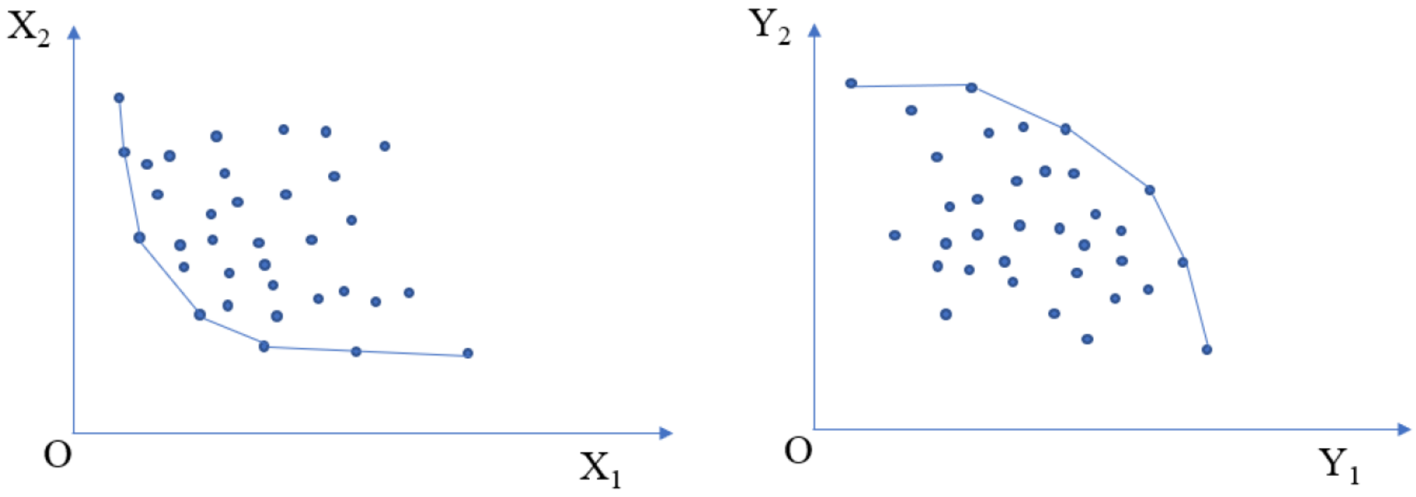
Estimation of **technology set**

- Parametric methods
  - describe the frontier in some specific functional form
  - use econometric methods to obtain the unknown parameters

- Nonparametric methods
  - use observed data to construct the frontier

Model

- Estimation of **technology set**



$$T = \left\{ (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{b}) : \sum_{j=1}^{J} \lambda_j \boldsymbol{x}_j \leq \boldsymbol{x}, \sum_{j=1}^{J} \lambda_j \boldsymbol{y}_j \geq \boldsymbol{y}, \sum_{j=1}^{J} \lambda_j \boldsymbol{b}_j = \boldsymbol{b}, \& \boldsymbol{\lambda} \geq 0 \right\}$$

Model

- Estimation of **radial DDF**

$$D_r(\boldsymbol{x},\boldsymbol{y},\boldsymbol{b};\boldsymbol{g}) = \sup\{\beta : ((\boldsymbol{x},\boldsymbol{y},\boldsymbol{b}) + \beta\boldsymbol{g}) \in T\}$$

$$\begin{aligned} D_r(\boldsymbol{x},\boldsymbol{y},\boldsymbol{b};\boldsymbol{g}) = & \max_{\beta,\boldsymbol{\lambda}} \beta \\ \text{s.t.} & \sum_{j=1}^{J} \lambda_j \boldsymbol{x}_j \leq \boldsymbol{x} + \beta \boldsymbol{g}_x, \\ & \sum_{j=1}^{J} \lambda_j \boldsymbol{y}_j \geq \boldsymbol{y} + \beta \boldsymbol{g}_y, \\ & \sum_{j=1}^{J} \lambda_j \boldsymbol{b}_j = \boldsymbol{b} + \beta \boldsymbol{g}_b, \\ & \lambda_j \geq 0, j = 1, \ldots, J. \end{aligned}$$

Model

- Estimation of **non-radial DDF** $D_{nr}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{b};\boldsymbol{g}) = \sup\{\boldsymbol{w}^T\boldsymbol{\beta} : ((\boldsymbol{x},\boldsymbol{y},\boldsymbol{b}) + diag(\boldsymbol{\beta})\cdot\boldsymbol{g}) \in T\}$

$$
\begin{aligned}
D_{nr}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{b};\boldsymbol{g}) &= \max_{\boldsymbol{\beta},\boldsymbol{\lambda}} \boldsymbol{w}^T\boldsymbol{\beta} \\
\text{s.t.} \sum_{j=1}^{J} \lambda_j \boldsymbol{x}_j &\leq \boldsymbol{x} + diag(\boldsymbol{\beta}_x)\cdot\boldsymbol{g}_x, \\
\sum_{j=1}^{J} \lambda_j \boldsymbol{y}_j &\geq \boldsymbol{y} + diag(\boldsymbol{\beta}_y)\cdot\boldsymbol{g}_y, \\
\sum_{j=1}^{J} \lambda_j \boldsymbol{b}_j &= \boldsymbol{b} + diag(\boldsymbol{\beta}_b)\cdot\boldsymbol{g}_b, \\
\boldsymbol{\beta} \geq 0; \lambda_j &\geq 0, j = 1, ..., J.
\end{aligned}
$$

Model

- **Malmquist–Luenberger productivity index** (Chung, 1997)

$$ML = \left[\frac{1 + D_r^s(\boldsymbol{x}^s, \boldsymbol{y}^s, \boldsymbol{b}^s; \boldsymbol{g})}{1 + D_r^s(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{b}^t; \boldsymbol{g})} \times \frac{1 + D_r^t(\boldsymbol{x}^s, \boldsymbol{y}^s, \boldsymbol{b}^s; \boldsymbol{g})}{1 + D_r^t(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{b}^t; \boldsymbol{g})}\right]^{1/2} \qquad \boldsymbol{g} = (\boldsymbol{0}, \boldsymbol{y}, -\boldsymbol{b})$$

$$MLEFFCH = \frac{1 + D_r^s(\boldsymbol{x}^s, \boldsymbol{y}^s, \boldsymbol{b}^s; \boldsymbol{g})}{1 + D_r^t(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{b}^t; \boldsymbol{g})}$$

$$MLTECH = \left[\frac{1 + D_r^t(\boldsymbol{x}^s, \boldsymbol{y}^s, \boldsymbol{b}^s; \boldsymbol{g})}{1 + D_r^s(\boldsymbol{x}^s, \boldsymbol{y}^s, \boldsymbol{b}^s; \boldsymbol{g})} \times \frac{1 + D_r^t(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{b}^t; \boldsymbol{g})}{1 + D_r^s(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{b}^t; \boldsymbol{g})}\right]^{1/2}$$

Model

- **Luenberger productivity indicator** (Färe and Grosskopf, 2010)

$$L = \left[\left(D_{nr}^{t}(\boldsymbol{x}^{s}, \boldsymbol{y}^{s}, \boldsymbol{b}^{s}; \boldsymbol{g}) - D_{nr}^{t}(\boldsymbol{x}^{t}, \boldsymbol{y}^{t}, \boldsymbol{b}^{t}; \boldsymbol{g}) \right] \times \frac{1}{2} + \left[D_{nr}^{s}(\boldsymbol{x}^{s}, \boldsymbol{y}^{s}, \boldsymbol{b}^{s}; \boldsymbol{g}) - D_{nr}^{s}(\boldsymbol{x}^{t}, \boldsymbol{y}^{t}, \boldsymbol{b}^{t}; \boldsymbol{g})) \right] \times \frac{1}{2}$$

$$\boldsymbol{g} = (\boldsymbol{0}, \boldsymbol{y}, -\boldsymbol{b})$$

$$LEFFCH = D_{nr}^{s}(\boldsymbol{x}^{s}, \boldsymbol{y}^{s}, \boldsymbol{b}^{s}; \boldsymbol{g}) - D_{nr}^{t}(\boldsymbol{x}^{t}, \boldsymbol{y}^{t}, \boldsymbol{b}^{t}; \boldsymbol{g})$$

$$LTECH = \left[D_{nr}^{t}(\boldsymbol{x}^{t}, \boldsymbol{y}^{t}, \boldsymbol{b}^{t}; \boldsymbol{g}) - D_{nr}^{s}(\boldsymbol{x}^{t}, \boldsymbol{y}^{t}, \boldsymbol{b}^{t}; \boldsymbol{g}) \right] \times \frac{1}{2} + \left[\left(D_{nr}^{t}(\boldsymbol{x}^{s}, \boldsymbol{y}^{s}, \boldsymbol{b}^{s}; \boldsymbol{g}) - D_{nr}^{s}(\boldsymbol{x}^{s}, \boldsymbol{y}^{s}, \boldsymbol{b}^{s}; \boldsymbol{g}) \right] \times \frac{1}{2}$$

Stata commands

- **teddf** estimates directional distance function with undesirable outputs.
  - radial Debreu-Farrell measures, non-radial Russell measures
  - different production technology, e.g., sequential, global
- **gtfpch** estimates total factor productivity change with undesirable outputs
  - Malmquist–Luenberger productivity index
  - Luenberger indicator

Stata commands

- The **teddf** command

teddf *Xvarlist* = *Yvarlist*:*Bvarlist* [*if*] [*in*], dmu(*varname*) [
time(*varname*) gx(*varlist*) gy(*varlist*) gb(*varlist*) nonradial
wmat(name) vrs window(#) biennial sequential global tol(real)
maxiter(#) saving(filename[,replace])]

Stata commands

- The **gtfpch** command

```
gtfpch Xvarlist = Yvarlist:Bvarlist [if] [in] , [ dmu(varname) luenberger
    ort(string) gx(varlist) gy(varlist) gb(varlist) nonradial wmat(name)
    window(#) biennial sequential global fgnz rd tol(real)
    maxiter(#) saving(filename[,replace]) ]
```

Illustrative Example

- a data set of China's provinces (Yan et al., 2020)

```
.
. use example.dta

.
. describe

Contains data from example.dta
  obs:              90
 vars:               7                  6 Aug 2020 12:12
-----------------------------------------------------------------------------------------------
              storage   display
variable name   type    format          variable label
-----------------------------------------------------------------------------------------------
Province        str12   %12s           province name
year            int     %10.0g         year
K               float   %9.0g          capital stock (in 100 million 1997 CNY)
L               double  %10.0g         employment (in 10 thousand persons)
E               double  %10.0g         energy consumption (in million tons of standard coal)
Y               float   %9.0g          real GDP (in 100 million 1997 CNY)
CO2             float   %15.1f         carbon dioxide emission (in kg)
-----------------------------------------------------------------------------------------------
```

Illustrative Example

- Estimation of **radial DDF** (Chung et al., 1997)

```
. teddf K L= Y: CO2, dmu( Province ) time(year) sav(ex1result,replace)
 The diectional vector is (-K -L Y -CO2)

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

	Row	Province	year	Dval
1.	1	Anhui	2013	0.2917
2.	2	Anhui	2014	0.3589
3.	3	Anhui	2015	0.3735
4.	4	Beijing	2013	-0.0000
5.	5	Beijing	2014	-0.0000
6.	6	Beijing	2015	-0.0000
7.	7	Chongqing	2013	0.2068
8.	8	Chongqing	2014	0.2362
9.	9	Chongqing	2015	0.2570
10.	10	Fujian	2013	0.0877
11.	11	Fujian	2014	0.1423
12.	12	Fujian	2015	0.1482
13.	13	Gansu	2013	0.2894
14.	14	Gansu	2014	0.3679
15.	15	Gansu	2015	0.4425
16.	16	Guangdong	2013	-0.0000
17.	17	Guangdong	2014	0.0372
18.	18	Guangdong	2015	0.0487

$$D_r(\boldsymbol{x},\boldsymbol{y},\boldsymbol{b};\boldsymbol{g}) = \max_{\beta,\boldsymbol{\lambda}} \beta$$

$$\text{s.t.} \sum_{j=1}^{J} \lambda_j \boldsymbol{x}_j \le \boldsymbol{x} + \beta \boldsymbol{g}_x,$$

$$\sum_{j=1}^{J} \lambda_j \boldsymbol{y}_j \ge \boldsymbol{y} + \beta \boldsymbol{g}_y,$$

$$\sum_{j=1}^{J} \lambda_j \boldsymbol{b}_j = \boldsymbol{b} + \beta \boldsymbol{g}_b,$$

$$\lambda_j \ge 0, j = 1, ..., J.$$

Illustrative Example

- Estimation of **non-radial DDF**

```
teddf K L= Y: CO2, dmu( Province ) time(year) nonr sav(ex2result,replace)
The weight vector is (1 1 1 1)
The diectional vector is (-K -L Y -CO2)

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

	Row	Province	year	Dval	B_K	B_L	B_Y	B_CO2
1.	1	Anhui	2013	1.6710	0.4594	0.7225	0.0000	0.4890
2.	2	Anhui	2014	1.7823	0.5293	0.7198	0.0000	0.5331
3.	3	Anhui	2015	1.8210	0.5827	0.7181	0.0000	0.5202
4.	4	Beijing	2013	0.0000	0.0000	0.0000	0.0000	0.0000
5.	5	Beijing	2014	0.0000	0.0000	0.0000	0.0000	0.0000
6.	6	Beijing	2015	0.0000	0.0000	0.0000	0.0000	0.0000
7.	7	Chongqing	2013	1.3031	0.4994	0.5887	0.0000	0.2149
8.	8	Chongqing	2014	1.3988	0.5415	0.5781	0.0000	0.2792
9.	9	Chongqing	2015	1.3936	0.5777	0.5661	0.0000	0.2499
10.	10	Fujian	2013	0.7968	0.3578	0.4363	0.0000	0.0026
11.	11	Fujian	2014	1.0092	0.4289	0.4426	0.0000	0.1377
12.	12	Fujian	2015	0.9997	0.4915	0.4581	0.0000	0.0500
13.	13	Gansu	2013	1.9927	0.5204	0.7853	0.0000	0.6869
14.	14	Gansu	2014	2.2088	0.0000	0.4725	1.4444	0.2920
15.	15	Gansu	2015	2.3532	0.0000	0.3980	1.7971	0.1580
16.	16	Guangdong	2013	0.0000	0.0000	0.0000	0.0000	0.0000
17.	17	Guangdong	2014	0.6215	0.1373	0.4425	0.0000	0.0417
18.	18	Guangdong	2015	0.6649	0.1980	0.4420	0.0000	0.0250
19.	19	Guangxi	2013	1.5334	0.4916	0.7061	0.0000	0.3357
20.	20	Guangxi	2014	1.6170	0.5515	0.7041	0.0000	0.3613

$$D_{nr}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{b}; \boldsymbol{g}) = \max_{\boldsymbol{\beta}, \boldsymbol{\lambda}} \boldsymbol{w}^T \boldsymbol{\beta}$$

$$\text{s.t.} \sum_{j=1}^{J} \lambda_j \boldsymbol{x}_j \leq \boldsymbol{x} + diag(\boldsymbol{\beta}_x) \cdot \boldsymbol{g}_x,$$

$$\sum_{j=1}^{J} \lambda_j \boldsymbol{y}_j \geq \boldsymbol{y} + diag(\boldsymbol{\beta}_y) \cdot \boldsymbol{g}_y,$$

$$\sum_{j=1}^{J} \lambda_j \boldsymbol{b}_j = \boldsymbol{b} + diag(\boldsymbol{\beta}_b) \cdot \boldsymbol{g}_b,$$

$$\boldsymbol{\beta} \geq 0; \lambda_j \geq 0, j = 1, ..., J.$$

Illustrative Example

- Estimation of **Malmquist–Luenberger productivity index**

```
. egen id=group(Province)

. xtset id year
       panel variable:  id (strongly balanced)
        time variable:  year, 2013 to 2015
                delta:  1 unit

. gtfpch K L= Y: CO2, dmu( Province ) global sav(ex3result,replace)
 The diectional vector is (0 0 Y -CO2)

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

	Row	Province	id	Pdwise	TFPCH	TECH	TECCH
1.	2	Anhui	1	2013_2014	0.9832	0.9179	1.0711
2.	3	Anhui	1	2014_2015	0.9853	0.9027	1.0916
3.	5	Beijing	2	2013_2014	1.0383	1.0000	1.0383
4.	6	Beijing	2	2014_2015	1.0620	1.0000	1.0620
5.	8	Chongqing	3	2013_2014	1.0029	0.9788	1.0246
6.	9	Chongqing	3	2014_2015	1.0476	0.9348	1.1207
7.	11	Fujian	4	2013_2014	0.9707	0.9248	1.0496
8.	12	Fujian	4	2014_2015	1.0327	0.9665	1.0685
9.	14	Gansu	5	2013_2014	0.9721	0.9011	1.0788
10.	15	Gansu	5	2014_2015	0.9791	0.8768	1.1167
11.	17	Guangdong	6	2013_2014	1.0221	0.9556	1.0695
12.	18	Guangdong	6	2014_2015	1.0175	0.9823	1.0358
13.	20	Guangxi	7	2013_2014	1.0076	0.9709	1.0378
14.	21	Guangxi	7	2014_2015	1.0750	0.9640	1.1152

Illustrative Example

- Estimation of **Luenberger productivity indicator**

```
. gtfpch K L= Y: CO2, dmu( Province ) nonr  global sav(ex4result,replace)
 The weight vector is (0 0 1 1)
 The diectional vector is (0 0 Y -CO2)

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

	Row	Province	id	Pdwise	TFPCH	TECH	TECCH
1.	2	Anhui	1	2013_2014	-0.0676	-0.2281	0.1605
2.	3	Anhui	1	2014_2015	0.0214	-0.0597	0.0811
3.	5	Beijing	2	2013_2014	0.0832	-0.0000	0.0832
4.	6	Beijing	2	2014_2015	0.1705	0.0000	0.1705
5.	8	Chongqing	3	2013_2014	0.0175	-0.0564	0.0738
6.	9	Chongqing	3	2014_2015	0.0178	-0.1079	0.1257
7.	11	Fujian	4	2013_2014	-0.0378	-0.0947	0.0569
8.	12	Fujian	4	2014_2015	0.0640	-0.0590	0.1230
9.	14	Gansu	5	2013_2014	-0.1748	-0.3039	0.1291
10.	15	Gansu	5	2014_2015	-0.1423	-0.2188	0.0765

Outlook

$$
\begin{aligned}
&\min_{\mathbf{x}} \ \text{or} \ \max_{\mathbf{x}} \ \mathbf{c}\mathbf{x}' \\
&\quad \text{such that } \mathbf{A}_{\text{EC}}\mathbf{x}' = \mathbf{b}_{\text{EC}} \\
&\qquad\qquad \mathbf{A}_{\text{IE}}\mathbf{x}' \leq \mathbf{b}_{\text{IE}} \\
&\qquad \mathbf{lowerbd} \leq \mathbf{x} \leq \mathbf{upperbd}
\end{aligned}
$$

- help mata linearprogram (Mehrotra, 1992)

Step 1: Initialization
Step 2: Definition of linear programming problem
Step 3: Perform optimization
Step 4: Display or obtain results

Outlook

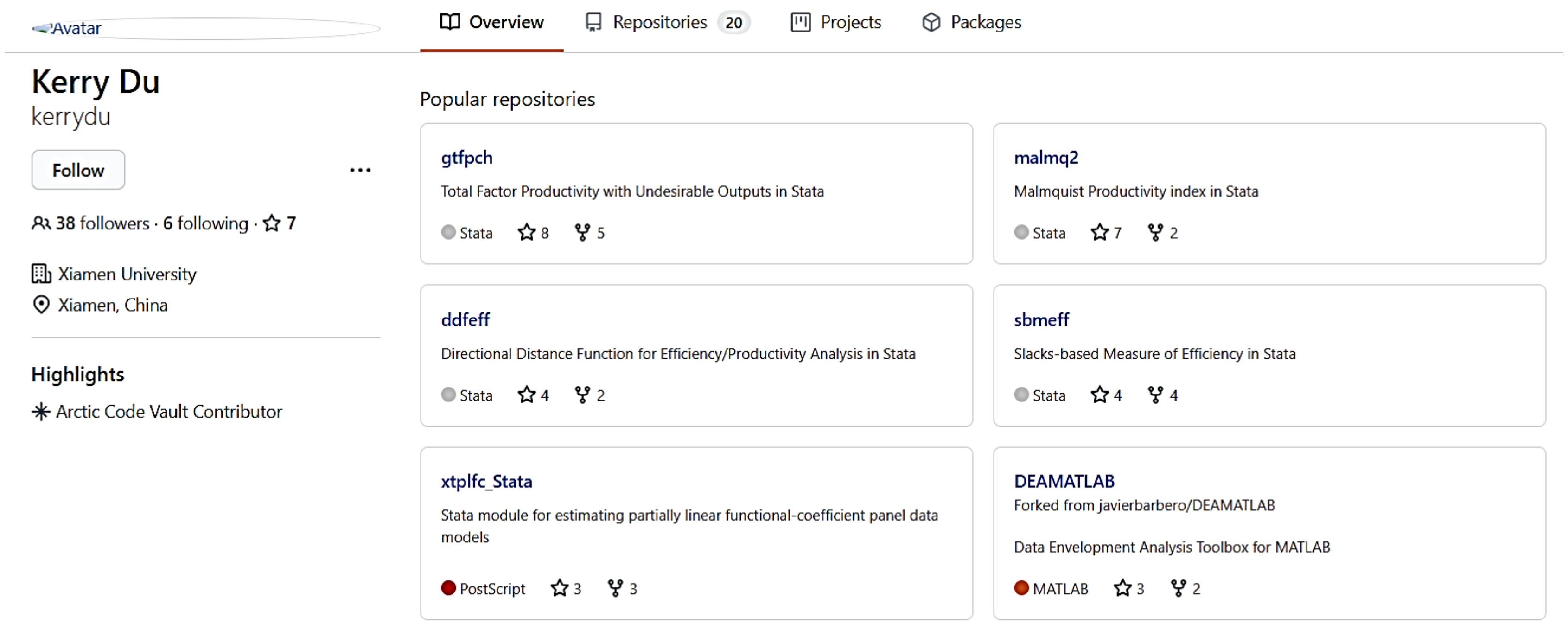
- help mata linearprogram (Mehrotra, 1992)

$$
\begin{aligned}
\max_{x_1, x_2} \quad & 5x_1 + 3x_2 \\
\text{such that} \quad & -x_1 + 11x_2 = 33 \\
& 0.5x_1 - x_2 \leq -3 \\
& 2x_1 + 14x_2 \leq 60 \\
& 2x_1 + x_2 \leq 14.5 \\
& x_1 - 0.4x_2 \leq 5 \\
& x_1 \geq 0 \\
& x_2 \geq 0
\end{aligned}
$$

```
mata:
    c = (5, 3)
    Aec = (-1, 11)
    bec = 33
    Aie = (0.5, -1 \ 2, 14 \ 2, 1 \ 1, -0.4)
    bie = (-3 \ 60 \ 14.5 \ 5)
    lowerbd = (0, 0)
    upperbd = (., .)
    q = LinearProgram()
    q.setCoefficients(c)
    q.setEquality(Aec, bec)
    q.setInequality(Aie, bie)
    q.setBounds(lowerbd, upperbd)
    q.optimize()
    q.parameters()
end
```

Outlook

- https://github.com/kerrydu/



- https://github.com/daopingw/

Thank You