Ekonomická fakulta
Faculty of Economics
Jihočeská univerzita v Českých Budějovicích
University of South Bohemia in České Budějovice

Data Envelopment Analysis within Evaluation of the Efficiency of Firm Productivity

Michal Houda

Department of Applied Mathematics and Informatics
Faculty of Economics, University of South Bohemia in České Budějovice

ROBUST 2018
Rybník, January 22, 2018

- *productivity growth*
  - output variations that are not explained by input variations
  - traditional approach: neglect inefficiencies in input/output usage
  - (total) productivity growth = shift in technologies
- Farrell, 1957
  - idea to measure productivity efficiency using all inputs (not only a selected one)
  - technical efficiency = multiplicative inverse of the Malmquist (1953) and Shephard (1957) input *distance function*
- Charnes, Cooper, Rhodes (1978)
  - successful attempt to compute productivity efficiency using linear optimization model
  - nonparametric approach
  - *data envelopment analysis* (DEA): the efficiency frontier made up as the boundary of a convex hull of the data points
  - different extensions to the model adopted

- DMU_k ... k-th decision making unit ($k = 1, \dots, K$)
- $X := (x_{ik}) \in \mathbb{R}^{m \times K}$ ... input matrix
  - $x_{\cdot k} := (x_{1k}, \dots, x_{mk})$ ... input vector of DMU_k
  - $x_{i\cdot} := (x_{i1}, \dots, x_{iK})$ ... values for i-th input ($i = 1, \dots, m$)
- $Y := (y_{jk}) \in \mathbb{R}^{n \times K}$ ... output matrix
  - $y_{\cdot k} := (y_{1k}, \dots, y_{nk})$ ... output vector of DMU_k
  - $y_{j\cdot} := (y_{j1}, \dots, y_{jK})$ ... values for j-th output ($j = 1, \dots, n$)
- PPS ... production possibility set – combination of allowed inputs and outputs
- DMU_0 with $(x_{\cdot 0}, y_{\cdot 0})$ ... DMU to be analyzed

Definition 1

DMU_1 dominates DMU_2 wrt. PPS if $x \leq x_{\cdot 0}$ and $y \geq y_{\cdot 0}$ with at least one (one-dimensional, input or output) inequality strict

Definition 2

DMU_0 is efficient wrt. PPS if $\nexists (x, y) \in \text{PPS}$ dominating $(x_{\cdot 0}, y_{\cdot 0})$.

- DMU_k ... k-th decision making unit ($k = 1, \dots, K$)
- $X := (x_{ik}) \in \mathbb{R}^{m \times K}$ ... input matrix
  - $x_{\cdot k} := (x_{1k}, \dots, x_{mk})$ ... input vector of DMU_k
  - $x_{i\cdot} := (x_{i1}, \dots, x_{iK})$ ... values for i-th input ($i = 1, \dots, m$)
- $Y := (y_{jk}) \in \mathbb{R}^{n \times K}$ ... output matrix
  - $y_{\cdot k} := (y_{1k}, \dots, y_{nk})$ ... output vector of DMU_k
  - $y_{j\cdot} := (y_{j1}, \dots, y_{jK})$ ... values for j-th output ($j = 1, \dots, n$)
- PPS ... production possibility set – combination of allowed inputs and outputs
- DMU_0 with $(x_{\cdot 0}, y_{\cdot 0})$ ... DMU to be analyzed

Definition 1

DMU_1 dominates DMU_2 wrt. PPS if $x \leq x_{\cdot 0}$ and $y \geq y_{\cdot 0}$ with at least one (one-dimensional, input or output) inequality strict

Definition 2

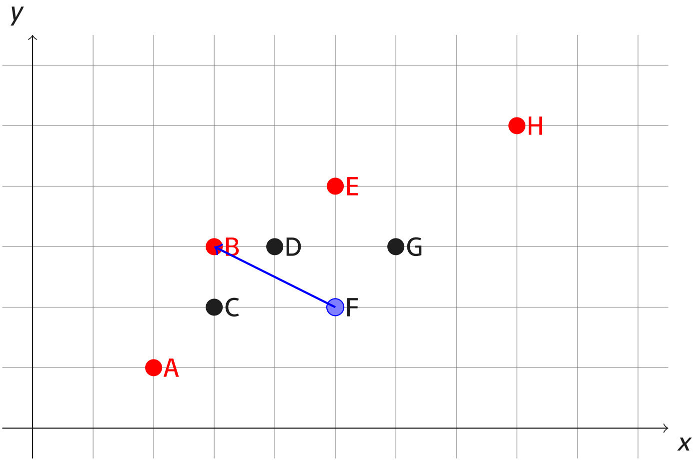
DMU_0 is efficient wrt. PPS if $\nexists (x, y) \in \text{PPS}$ dominating $(x_{\cdot 0}, y_{\cdot 0})$.

Discrete PPS (Bowlin, Brennan et al, 1984): $\text{PPS}_I = \{(x_k, y_k)\}_{k=1}^{K}$

Dominance wrt. PPS_I: additive model with integer constraints

$$
\begin{aligned}
&\max\left(\sum_j s_j^+ + \sum_i s_i^-\right) \text{ subject to} \\
&\qquad \sum_k x_{ik}\lambda_k + s_i^- = x_{i0} \qquad \forall i \quad \text{(inputs)} \\
&\qquad \sum_k y_{ik}\lambda_k - s_i^+ = y_{j0} \qquad \forall j \quad \text{(outputs)} \\
&\sum_k \lambda_k = 1,\ \lambda_k \in \{0,1\}^K,\ s_i^-, s_j^+ \geq 0
\end{aligned}
\tag{1}
$$

- s^- ... slack for $X\lambda \leq x_0$ // s^+ ... slack (surplus) for $Y\lambda \geq y_0$
- DMU_0 is efficient wrt. PPS_I if no slack is greater than 0 (i. e., both inequalities are active) in optimal solution



y
x
A
B
C
D
E
F
G
H

Continuous (convex) PPS (Banker, Cooper, Charnes, 1984):

$$\mathrm{PPS}_C = \{(x,y) \mid x = X\lambda, y = Y\lambda, \sum \lambda_k = 1, \lambda \geq 0\}$$

Dominance wrt. PPS_C: BCC-I model

$$\min \theta + \epsilon\left(\sum_j s_j^+ + \sum_i s_i^-\right) \text{ subject to}$$
$$\sum_k x_{ik}\lambda_k + s_i^- = \theta x_{i0} \qquad \forall i \quad \text{(inputs)}$$
$$\sum_k y_{ik}\lambda_k - s_i^+ = y_{j0} \qquad \forall j \quad \text{(outputs)} \tag{2}$$
$$\sum_k \lambda_k = 1,\ \lambda_k \geq 0, s_i^-, s_j^+ \geq 0, \theta \text{ unconstrained}$$

ϵ ... non-Archimedean infinitesimal

Dual problem:

$$\begin{aligned} \max v^T y_{.0} + q \text{ subject to} & \\ -u^T x_{.k} + v^T y_{.k} + q \leq 0 & \quad \forall k \quad \text{(DMUs)} \\ u^T x_{.0} = 1 & \quad \text{(dual for } \theta\text{)} \\ u \geq \epsilon \mathbf{1},\ v \geq \epsilon \mathbf{1},\ q \text{ unconstrained} & \end{aligned} \tag{3}$$

q (dual for $\sum_k \lambda_k = 1$) ... variable returns to scale (VRS) factor

BCC-I DEA problem of fractional programming:

$$\begin{aligned} \max \frac{v^T y_{.0} + q}{u^T x_{.0}} & \text{ subject to} \\ \frac{v^T y_{.k} + q}{u^T x_{.k}} \leq 1 & \quad \forall k \quad \text{(DMUs)} \\ u^T x_{.0} = 1, u/u^T x_{.0} \geq \epsilon \mathbf{1},\ v/u^T x_{.0} \geq \epsilon \mathbf{1},\ q \text{ unconstrained} & \end{aligned} \tag{4}$$

Dual problem:

$$\begin{aligned} \max v^T y_{\cdot 0} + q \text{ subject to} & \\ -u^T x_{\cdot k} + v^T y_{\cdot k} + q \leq 0 & \quad \forall k \quad \text{(DMUs)} \\ u^T x_{\cdot 0} = 1 & \quad \text{(dual for } \theta) \\ u \geq \epsilon \mathbf{1},\ v \geq \epsilon \mathbf{1},\ q \text{ unconstrained} & \end{aligned} \tag{3}$$

q (dual for $\sum_k \lambda_k = 1$) ... variable returns to scale (VRS) factor

BCC-I DEA problem of fractional programming:

$$\begin{aligned} \max \frac{v^T y_{\cdot 0} + q}{u^T x_{\cdot 0}} & \text{ subject to} \\ \frac{v^T y_{\cdot k} + q}{u^T x_{\cdot k}} \leq 1 & \quad \forall k \quad \text{(DMUs)} \\ u^T x_{\cdot 0} = 1, u/u^T x_{\cdot 0} \geq \epsilon \mathbf{1},\ v/u^T x_{\cdot 0} \geq \epsilon \mathbf{1},\ & q \text{ unconstrained} \end{aligned} \tag{4}$$

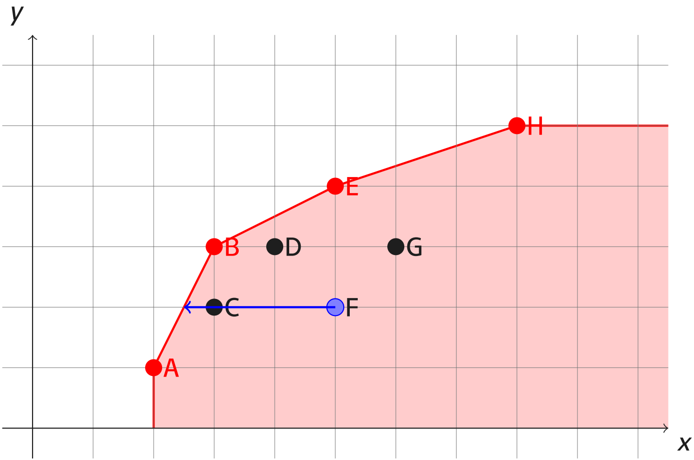
Definition 3 (DEA Efficiency)

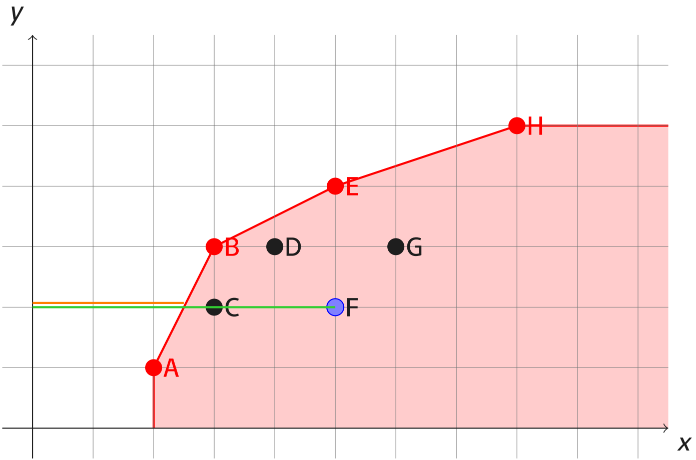
DMU_0 is BCC-I (fully) efficient wrt. PPS_C if

1. $\theta^* = 1$
2. $s^{+*} = s^{-*} = 0$

Remark

- weak DEA efficiency: $\theta^* = 1$ but some of s_i^{-*}, s_j^{+*} are not zero (efficient points which are not extreme points of PPS)
- two-stage solution procedure:
  1. solve the BCC-I problem with $\epsilon = 0$ to obtain θ^*
  2. solve the problem $\max \sum_j s_j^+ + \sum_i s_i^-$ subject to remaining constraints where $\epsilon = 0$ and $\theta = \theta^*$ to obtain maximal possible slacks


y
x
A
B
C
D
E
F
G
H



y
H
E
B
D
G
C
F
A
x

Linear PPS (Charnes, Cooper, Rhodes (1978)):

$$\mathrm{PPS}_L = \left\{(x, y) \mid x = X\lambda, y = Y\lambda, \lambda \geq 0\right\}$$

Dominance wrt. PPS_L: CCR-I model

$$\begin{aligned} \min \theta + \epsilon\Big(\sum_j s_j^+ + \sum_i s_i^-\Big) \text{ subject to} \\ \sum_k x_{ik}\lambda_k + s_i^- &= \theta x_{i0} \quad \forall i \quad \text{(inputs)} \\ \sum_k y_{ik}\lambda_k - s_i^+ &= y_{j0} \quad \forall j \quad \text{(outputs)} \\ \lambda_k \geq 0, s_i^-, s_j^+ \geq 0, \theta \text{ unconstrained} \end{aligned} \tag{5}$$

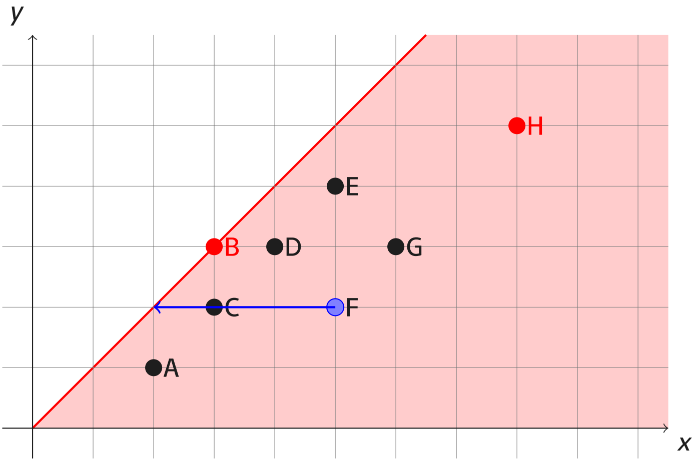
Dual problem:

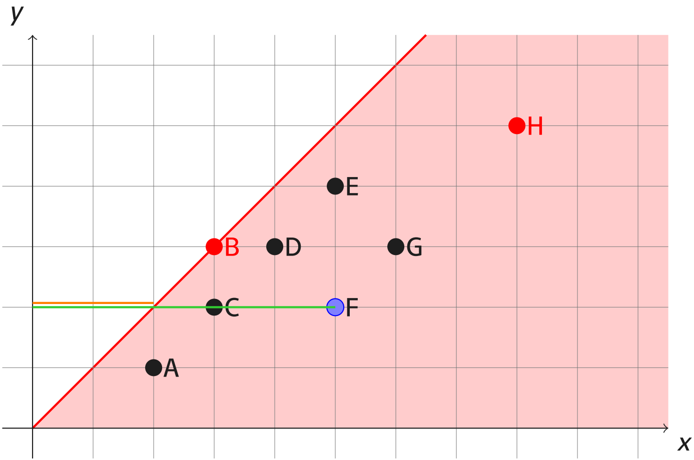
$$\begin{aligned} \max v^T y_{\cdot 0} \text{ subject to} & \\ -u^T x_{\cdot k} + v^T y_{\cdot k} \leq 0 & \quad \forall k \quad \text{(DMUs)} \\ u^T x_{\cdot 0} = 1 & \quad \text{dual for } \theta) \\ u \geq \epsilon \mathbf{1}, \ v \geq \epsilon \mathbf{1} & \end{aligned} \tag{6}$$

$q = 0$ ... constant returns to scale (CRS)

CCR-I DEA problem of fractional programming:

$$\begin{aligned} \max \frac{v^T y_{\cdot 0}}{u^T x_{\cdot 0}} \text{ subject to} & \\ \frac{v^T y_{\cdot k}}{u^T x_{\cdot k}} \geq 1 & \quad \forall k \quad \text{(DMUs)} \\ u^T x_{\cdot 0} = 1, u/u^T x_{\cdot 0} \geq \epsilon \mathbf{1}, \ v/u^T x_{\cdot 0} \geq \epsilon \mathbf{1} & \end{aligned} \tag{7}$$


y
x
H
E
B
D
G
C
F
A



y
x
H
E
B
D
G
C
F
A

Directional Distance Models (Chambers, Chung, Färe (1996, 1998)):

- dealing with negative data
- g^x, g^y ... vectors of improvement directions

Generic Directional Distance Model (wrt. PPS_C):

$$\begin{aligned} &\max \beta \text{ subject to} \\ &\quad \sum_k x_{ik}\lambda_k \le x_{i0} - \beta g_i^x \qquad \forall i \quad \text{(inputs)} \\ &\quad \sum_k y_{ik}\lambda_k \ge y_{j0} + \beta g_j^y \qquad \forall j \quad \text{(outputs)} \\ &\quad \sum_k \lambda_k = 1,\ \lambda_k \ge 0,\ \beta_0 \ge 0 \end{aligned} \tag{8}$$

- efficiency of DMU_0: $\beta^* = 0$
- special case: $g^x = x_0$, $g^y = 0$, $\theta = 1 - \beta$: BCC-I case

Range Directional Model

- range of possible improvements:

$$g_i^x = x_{i0} - \min_k x_{ik}$$
$$g_j^y = \max_k y_{jk} - y_{j0}$$

$I = (\min_k x_{\cdot k}, \max_k y_{\cdot k})$ ... ideal point

$$\max \beta \text{ subject to}$$
$$\sum_k x_{ik}\lambda_k \leq (1-\beta)x_{i0} + \beta \min_k x_{ik} \quad \forall i \quad \text{(inputs)}$$
$$\sum_k y_{ik}\lambda_k \geq (1-\beta)y_{j0} + \beta \min_k y_{jk} \quad \forall j \quad \text{(outputs)} \tag{9}$$
$$\sum_k \lambda_k = 1, \ \lambda_k \geq 0, \ \beta_0 \geq 0$$

Production possibility sets (available technology):

$$\text{PPS}_L := \{(x,y) \mid x = X\lambda, y = Y\lambda, \lambda \geq 0\}$$

$$\text{PPS}_C := \{(x,y) \mid x = X\lambda, y = Y\lambda, \sum \lambda_k = 1, \lambda \geq 0\}$$

$$\text{PPS}_C^s := \{(x,y) \mid x = X\lambda + s^+, y = Y\lambda - s^-, \sum \lambda_k = 1, \lambda \geq 0, s^+, s^- \geq 0\}$$

$$\text{PPS}_I := \{(x,y) \mid x = X\lambda, y = Y\lambda, \sum \lambda_k = 1, \lambda \in \{0,1\}^K\}$$

$$\text{PPS} := \{(x,y) \mid y \text{ can be produced from } x\} \qquad \text{(general PPS)}$$

Desirable properties for PPS

1. convexity: if $(x_{\cdot k}, y_{\cdot k}) \in$ PPS and $\lambda \geq 0, \sum \lambda_k = 1$ then $(X\lambda, Y\lambda) \in$ PPS
2. free (strong) disposability of inputs and outputs:
    1. if $(x,y) \in$ PPS and $x^+ := x + s^+$ with $s^+ \geq 0$ then $(x^+, y) \in$ PPS
    2. if $(x,y) \in$ PPS and $y^- := y - s^-$ with $s^- \geq 0$ then $(x, y^-) \in$ PPS
3. minimum intersection: PPS is the intersection of all sets $\widehat{\text{PPS}}$ satisfying properties 1 and 2, subject to $(x,y) \in \widehat{\text{PPS}}$

Production possibility sets (available technology):

$$\text{PPS}_L := \{(x,y) \mid x = X\lambda, y = Y\lambda, \lambda \geq 0\}$$

$$\text{PPS}_C := \{(x,y) \mid x = X\lambda, y = Y\lambda, \sum \lambda_k = 1, \lambda \geq 0\}$$

$$\text{PPS}_C^s := \{(x,y) \mid x = X\lambda + s^+, y = Y\lambda - s^-, \sum \lambda_k = 1, \lambda \geq 0, s^+, s^- \geq 0\}$$

$$\text{PPS}_I := \{(x,y) \mid x = X\lambda, y = Y\lambda, \sum \lambda_k = 1, \lambda \in \{0,1\}^K\}$$

$$\text{PPS} := \{(x,y) \mid y \text{ can be produced from } x\} \qquad \text{(general PPS)}$$

Additional desirable properties for PPS

4. no free lunch: if $(0, y) \in$ PPS then $y = 0$
5. no infinite outputs: $A(x) := \{(u, y) \mid u \leq x\}$ is bounded $\forall x$
6. closeness: PPS is closed (technical property)

Usual assumption (may be eliminated by some extensions)

7. no negative inputs and outputs

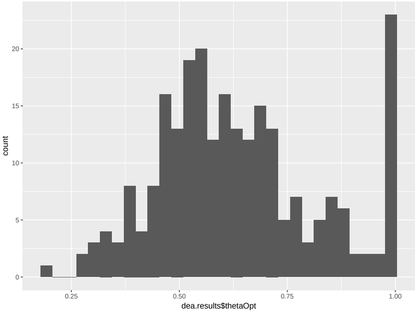
- Data: annual accounts of 380 Czech companies from the food industry (NACE C.10) [selected year: 2014]
- Implementation:
  - grouping the companies (according to the EC classification of economic activities)
  - choosing appropriate inputs and outputs to be analysed
  - choosing the model (returns to scale)
  - computer implementation
- Issues:
  - missing or implausible data
  - negative inputs/outputs

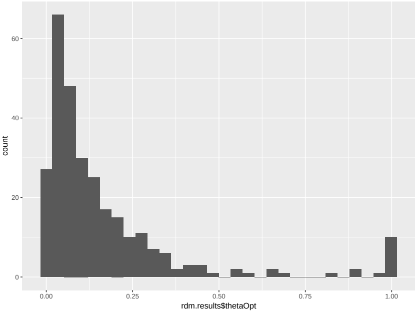
- Companies:
  - the whole group C.10 (Manufacture of food products)
- Inputs
  - `SPMAAEN`: material and energy consumption – 89 companies with no costs reported
  - `ON`: personnel costs
  - `STALAA`: fixed assets (buildings, equipments)
  - `POSN`: percentage of the personnel costs
- Outputs
  - `VYKONY`: business performance
  - `ROA`: return on assets (earning before interest and taxes per total assets) – 70 companies having negative ROA
- Models and feasible solutions:
  - input-oriented with variable returns to scale: 244 companies
  - range directional model: 291 companies

Considered alternatives (not in today's presentation):

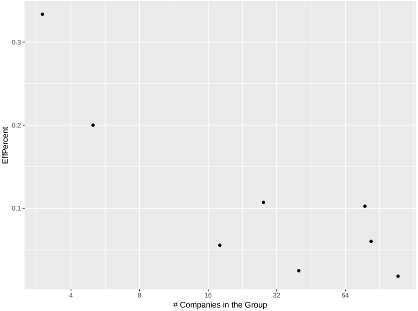
- Groups of companies:
  - manufacture and processing of meat (78) / fish (3) /fruit and vegetables (18) / oils and fats (5) / dairy produts (28) / grain mills products (15) / bakery and farinaceous products (109) / other food products (83) / prepared animal feeds (40)
- Inputs:
  - production consumption / depreciations / tangible and intangible fixed assets / cost of capital
- Outputs:
  - value of sales of goods and services / operating income / EBIT (earnings before interest and taxes) / value added
- Models: Cooper, Seiford, Tone (2007), Cooper, Seiford, Zhoe (2011):
  - input/output oriented CRS/VRS DEA models with discretionary/non-discretionary ou
  - alternative DEA models: additive (translation invariant), slack-based, Russel, free disposal hull, other directional distance

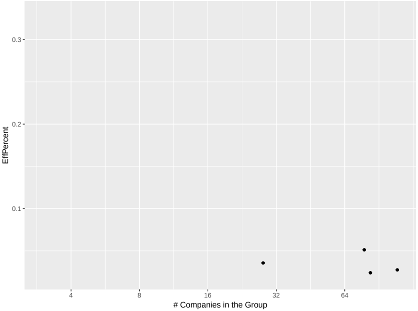
- BCC: 22 efficient companies (additional 3 with efficiency $> 95\,\%$)
- RDM: 10 efficient companies



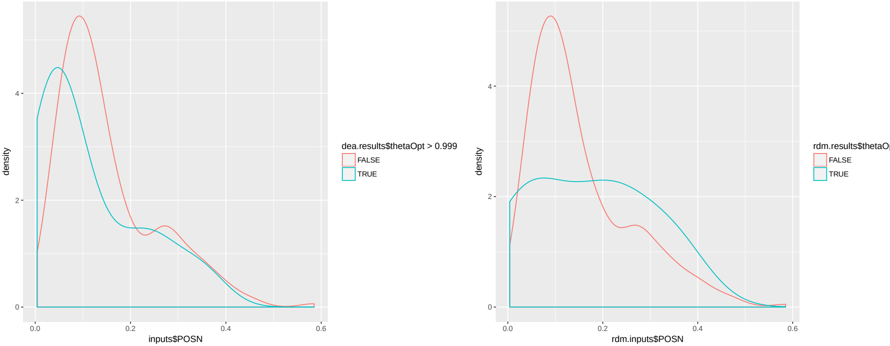


- BCC: from the groups, manufacture and processing of meat (8 of 78), manufacture of other food products (5 of 83)

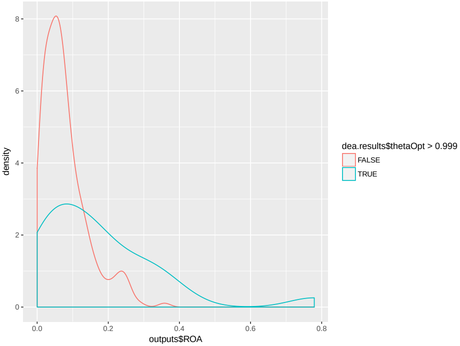


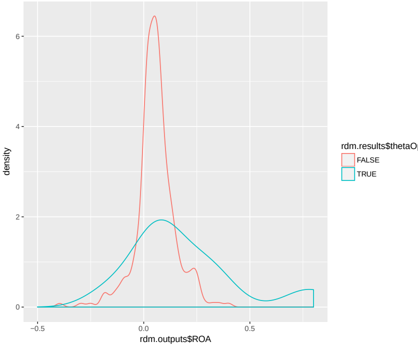


- distribution of a selected input for efficient and inefficient companies

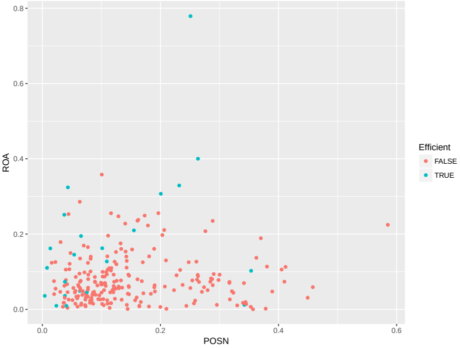


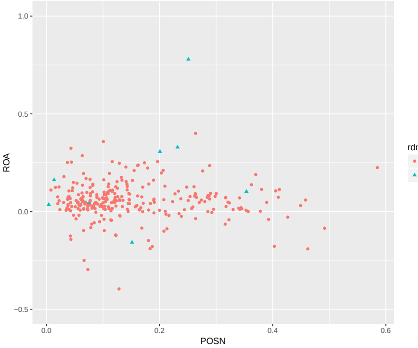
- distribution of a selected output for efficient and inefficient companies





- distribution of a selected input and a selected output for efficient and inefficient companies





- Caves, Christensen, Diewert (1982)
  - input-based Malmquist productivity index defined as the ratio of two input distance functions (optimal values of DEA problems)
- Färe, Grosskopf, Lindgren, Roos (1992)
  - introducing dynamics: the Malmquist index defined as the geometric mean of two indexes in Caves et al.'s sense (four DEA problems computed)
  - the index can be decomposed into two components: an *efficiency change* (the ratio of the technical efficiencies in two time periods), and a *technical change* (the shift of the frontier between two time periods)
  - input-oriented DEA model with CRS used to calculate the input distances
- further studies: using different distance functions / DEA models to calculate the index, e. g.
  - Chung, Färe, Grosskopf (1997), Oh (2010): (local and global) Malmquist-Luenberger index (using directional distance)
  - Asmild, Baležentis, Hougaard (2016): multi-directional efficiency
  - Boussemart, Briec et al. (2009): (generalized) α-returns to scale

- define an *input (output) distance function* $D(x_0, y_0)$ for the DMU_0 as the inverse of the optimal value of the input (output) based DEA problem under technology PPS:

$$D(x_0, y_0) := \frac{1}{\theta^*}$$

- take these distance for two different time periods t, $t+1$
- define the *Malmquist index* as

$$M_t^{t+1} := \sqrt{\frac{D^t(x_0^{t+1}, y_0^{t+1})}{D^t(x_0^t, y_0^t)} \cdot \frac{D^{t+1}(x_0^{t+1}, y_0^{t+1})}{D^{t+1}(x^t, y^t)}}$$

- two components of the Malmquist index:

$$M_t^{t+1} := EC_t^{t+1} \cdot TC_t^{t+1}$$

  - efficiency change $EC_t^{t+1} := \dfrac{D^t(x_0^{t+1}, y_0^{t+1})}{D^t(x_0^t, y_0^t)}$
  - technology change $TC_t^{t+1} := \sqrt{\dfrac{D^t(x_0^{t+1}, y_0^{t+1})}{D^{t+1}(x_0^{t+1}, y_0^{t+1})} \cdot \dfrac{D^t(x^t, y^t)}{D^{t+1}(x^t, y^t)}}$

References I

Cooper, W. W., Seiford, L. M. and Zhu, J., eds. (2011).
Handbook on Data Envelopment Analysis.
Springer, New York, 2nd ed.

Cooper, W. W., Seiford, L. M. and Tone, K., eds. (2007).
Data Envelopment Analysis.
Springer, New York, 2nd ed.

Shephard, R. W. (1953).
Cost and Production Functions.
Princeton University Press, Princeton.

References II

Malmquist, S. (1953).
Index numbers and indifference surfaces
Trabajos de Estadistica, **4**(2), 209–242.

Farrel, M. J. (1957)
The Measurement of Productive Efficiency
Journal of the Royal Statistical Society. Series A (General), **120**(3), 253–290.

Charnes, A., Cooper, W. W., Rhodes, E. (1978)
Measuring the efficiency of decision making units.
European Journal of Operational Research, **2**(6), 429–444

Banker, R. D. and Charnes, A. and Cooper, W. W. (1984)
Some models for estimating technical and scale inefficiencies in data envelopment analysis
Management Science, **30**(9), 1078–1092.

Bowlin, W. F., Brennan, J., Cooper, W. W. and Sueyoshi, T. (1984)
DEA models for evaluating efficiency dominance
Research Report of the University of Texas, Center for Cybernetic Studies, Austin

Chambers, R. G., Chung, Y. and Färe, R. (1996)
Benefit and distance functions.
Journal of Economic Theory, **70**(2), 407–419.

References III

Caves, D. W., Christensen, L. R., Diewert, W. E. (1982)
The economic theory of index numbers and the measurement of input, output, and productivity
Econometrica, **50**(6), 1393–1414

Färe, R., Grosskopf, S., Lindgren, B., Roos, P. (1992)
Productivity changes in Swedish pharmacies 1980–1989: A non-parametric Malmquist approach
Journal of Productivity Analysis, **3**, 85–101

Chung, Y. H., Färe, R., Grosskopf, S. (1997)
Productivity and undesirable outputs: a directional distance function approach
Journal of Environmental Management, **51**(3), 229-240

Oh, D.-H. (2010)
A global Malmquist-Luenberger productivity index
Journal of Productivity Analysis,**34**(3), 183-197

Boussemart, J.-P., Briec, W., Peypoch, N., Tavéra, C. (2009)
α-returns to scale and multi-output production technologies
European Journal of Productivity Analysis, **197**(1), 332–339

Asmild, M., Baležentis, T., Hougaard, J. L. (2016)
Multi-directional productivity change: MEA-Malmquist
Journal of Productivity Analysis, **46**(2–3), 109–119