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
**Introduction**

**Measuring Productivity Efficiency**

- **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 \((k = 1, \ldots, K)\)**
- **\(X := (x_{ik}) \in \mathbb{R}^{m \times K}\) ... input matrix**
  - **\(x_{.k} := (x_{1k}, \ldots, x_{mk})\) ... input vector of DMU\(_k\)**
  - **\(x_{i.} := (x_{i1}, \ldots, x_{iK})\) ... values for \(i\)-th input \((i = 1, \ldots, m)\)**
- **\(Y := (y_{jk}) \in \mathbb{R}^{n \times K}\) ... output matrix**
  - **\(y_{.k} := (y_{1k}, \ldots, y_{nk})\) ... output vector of DMU\(_k\)**
  - **\(y_{j.} := (y_{j1}, \ldots, y_{jK})\) ... values for \(j\)-th output \((j = 1, \ldots, n)\)**
- **PPS ... production possibility set – combination of allowed inputs and outputs**
- **DMU\(_0\) with \((x_{.0}, y_{.0})\) ... DMU to be analyzed**

### Definition 1

DMU\(_1\) dominates DMU\(_2\) wrt. PPS if \(x \leq x_{.0}\) and \(y \geq y_{.0}\) 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 \((x_{.0}, y_{.0})\).
Data Envelopment Analysis
Notation, Efficiency Dominance

- DMU<sub>k</sub> ... k-th decision making unit (<i>k = 1, \ldots, K</i>)
- \( X := (x_{ik}) \in \mathbb{R}^{m \times K} \) ... input matrix
  - \( x_k := (x_{1k}, \ldots, x_{mk}) \) ... input vector of DMU<sub>k</sub>
  - \( x_i := (x_{i1}, \ldots, x_{ik}) \) ... values for \( i \)-th input (<i>i = 1, \ldots, m</i>)
- \( Y := (y_{jk}) \in \mathbb{R}^{n \times K} \) ... output matrix
  - \( y_k := (y_{1k}, \ldots, y_{nk}) \) ... output vector of DMU<sub>k</sub>
  - \( y_j := (y_{j1}, \ldots, y_{jk}) \) ... values for \( j \)-th output (<i>j = 1, \ldots, n</i>)
- PPS ... production possibility set – combination of allowed inputs and outputs
- DMU<sub>0</sub> with \( (x_0, y_0) \) ... DMU to be analyzed

**Definition 1**

DMU<sub>1</sub> dominates DMU<sub>2</sub> wrt. PPS if \( x \leq x_0 \) and \( y \geq y_0 \) with at least one (one-dimensional, input or output) inequality strict

**Definition 2**

DMU<sub>0</sub> is efficient wrt. PPS if \( \#(x, y) \in \text{PPS dominating} \ (x_0, y_0) \).
**Discrete PPS** (Bowlin, Brennan et al, 1984): \( \text{PPS}_I = \{(x_k, y_k)\}_{k=1}^K \)

Dominance wrt. \( \text{PPS}_I \): additive model with integer constraints

\[
\max \left( \sum_j s^+_j + \sum_i s^-_i \right) \quad \text{subject to}
\]

\[
\sum_k x_{ik} \lambda_k + s^-_i = x_i^0 \quad \forall i \quad \text{(inputs)}
\]

\[
\sum_k y_{ik} \lambda_k - s^+_i = y_j^0 \quad \forall j \quad \text{(outputs)}
\]

\[
\sum_k \lambda_k = 1, \quad \lambda_k \in \{0, 1\}^K, \quad s^-_i, s^+_j \geq 0
\]

- \( s^- \) ... slack for \( X \lambda \leq x_0 \)  //  \( s^+ \) ... slack (surplus) for \( Y \lambda \geq y_0 \)

- DMU\(_0\) is efficient wrt. \( \text{PPS}_I \) if no slack is greater than 0 (i.e., both inequalities are active) in optimal solution
Data Envelopment Analysis – 0-1 Model

Discrete Production Possibility Set
Data Envelopment Analysis – BCC Model
Continuous Production Possibility Set

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

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

Dominance wrt. PPS\(_C\): **BCC-I model**

\[
\begin{align*}
\min & \quad \theta + \epsilon \left( \sum_j s_j^+ + \sum_i s_i^- \right) \quad \text{subject to} \\
\sum_k x_{ik}\lambda_k + s_i^- &= \theta x_{i0} \quad \forall i \quad \text{(inputs)} \\
\sum_k y_{jk}\lambda_k - s_j^+ &= y_{j0} \quad \forall j \quad \text{(outputs)} \\
\sum \lambda_k &= 1, \lambda_k \geq 0, s_i^-, s_j^+ \geq 0, \theta \text{ unconstrained} \\
\epsilon &\quad \text{... non-Archimedean infinitesimal}
\end{align*}
\]
Data Envelopment Analysis – BCC Model

Continuous Production Possibility Set – Dual Problem

**Dual problem:**

\[
\max \, \nu^T y_0 + q \quad \text{subject to} \\
- u^T x_k + \nu^T y_k + q \leq 0 \quad \forall k \quad \text{(DMUs)} \\
\nu^T y_0 = 1 \quad \text{(dual for } \theta) \\
u \geq \epsilon 1, \, \nu \geq \epsilon 1, \, q \text{ unconstrained}
\]

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

**BCC-I DEA problem of fractional programming:**

\[
\max \, \frac{\nu^T y_0 + q}{u^T x_0} \quad \text{subject to} \\
\frac{\nu^T y_k + q}{u^T x_k} \leq 1 \quad \forall k \quad \text{(DMUs)} \\
u^T x_0 = 1, \, u/u^T x_0 \geq \epsilon 1, \, v/u^T x_0 \geq \epsilon 1, \, q \text{ unconstrained}
\]
Dual problem:

\[
\begin{align*}
\text{max } & \mathbf{v}^T \mathbf{y} \cdot 0 + q \text{ subject to } \\
& -u^T \mathbf{x} \cdot k + \mathbf{v}^T \mathbf{y} \cdot k + q \leq 0 \quad \forall k \quad \text{(DMUs)} \\
& u^T \mathbf{x} \cdot 0 = 1 \quad \text{(dual for } \theta) \\
& u \geq \epsilon \mathbf{1}, \ v \geq \epsilon \mathbf{1}, \ q \text{ unconstrained}
\end{align*}
\]

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

BCC-I DEA problem of fractional programming:

\[
\begin{align*}
\text{max } & \frac{\mathbf{v}^T \mathbf{y} \cdot 0 + q}{u^T \mathbf{x} \cdot 0} \text{ subject to } \\
& \frac{\mathbf{v}^T \mathbf{y} \cdot k + q}{u^T \mathbf{x} \cdot k} \leq 1 \quad \forall k \quad \text{(DMUs)} \\
& u^T \mathbf{x} \cdot 0 = 1, \ u / u^T \mathbf{x} \cdot 0 \geq \epsilon \mathbf{1}, \ v / u^T \mathbf{x} \cdot 0 \geq \epsilon \mathbf{1}, \ q \text{ unconstrained}
\end{align*}
\]
### Definition 3 (DEA Efficiency)

DMU₀ is **BCC-I (fully) efficient** wrt. PPSₐ 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 \( \theta^* \)
  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
Data Envelopment Analysis – BCC Model

Continuous Production Possibility Set
Data Envelopment Analysis – BCC Model
Continuous Production Possibility Set
Linear PPS (Charnes, Cooper, Rhodes (1978)):

\[ PPS_L = \{ (x, y) \mid x = X\lambda, y = Y\lambda, \lambda \geq 0 \} \]

Dominance wrt. \( PPS_L \): CCR-I model

\[
\begin{align*}
\min \theta + \epsilon (\sum_j s_j^+ + \sum_i s_i^-) & \quad \text{subject to} \\
\sum_k x_{ik} \lambda_k + s_i^- & = \theta x_{i0} \quad \forall i \quad \text{(inputs)} \\
\sum_k y_{ik} \lambda_k - s_i^+ & = y_{j0} \quad \forall j \quad \text{(outputs)} \\
\lambda_k & \geq 0, s_i^-, s_j^+ \geq 0, \theta \text{ unconstrained}
\end{align*}
\]
Data Envelopment Analysis – CCR Model
Linear Production Possibility Set – Dual Problem

Dual problem:

\[
\begin{align*}
\text{max } v^T y & \text{ subject to } \\
-u^T x_k + v^T y_k & \leq 0 \quad \forall k \quad \text{(DMUs)} \\
u^T x_0 &= 1 \quad \text{dual for } \theta) \\
& u \geq \epsilon 1, \ v \geq \epsilon 1
\end{align*}
\]

\[q = 0 \ldots \text{ constant returns to scale (CRS)}\]

CCR-I DEA problem of fractional programming:

\[
\begin{align*}
\text{max } \frac{v^T y}{u^T x} & \text{ subject to } \\
\frac{v^T y_k}{u^T x_k} & \geq 1 \quad \forall k \quad \text{(DMUs)} \\
u^T x_0 &= 1, \ u/u^T x_0 \geq \epsilon 1, \ v/u^T x_0 \geq \epsilon 1
\end{align*}
\]
Data Envelopment Analysis – CCR Model
Linear Production Possibility Set

\[ y \]

\[ x \]

Michal Houda
DEA within Firm Productivity
Data Envelopment Analysis – CCR Model

Linear Production Possibility Set

Michal Houda

DEA within Firm Productivity
**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\(_C\)):

\[
\begin{align*}
\max \beta & \quad \text{subject to} \\
\sum_k x_{ik} \lambda_k & \leq x_{i0} - \beta g^x_i \quad \forall i \quad \text{(inputs)} \\
\sum_k y_{ik} \lambda_k & \geq y_{j0} + \beta g^y_j \quad \forall j \quad \text{(outputs)} \\
\sum_k \lambda_k & = 1, \quad \lambda_k \geq 0, \quad \beta_0 \geq 0
\end{align*}
\]

- 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:

\[ g^x_i = x_{i0} - \min_k x_{ik} \]
\[ g^y_j = \max_k y_{jk} - y_{j0} \]

\[ l = (\min_k x_k, \max_k y_k) \] ... ideal point

\[
\begin{align*}
\max \beta & \quad \text{subject to} \\
\sum_k x_{ik} \lambda_k & \leq (1 - \beta) x_{i0} + \beta \min_k x_{ik} \quad \forall i \quad \text{(inputs)} \\
\sum_k y_{jk} \lambda_k & \geq (1 - \beta) y_{j0} + \beta \min_k y_{jk} \quad \forall j \quad \text{(outputs)} \\
\sum_k \lambda_k & = 1, \quad \lambda_k \geq 0, \quad \beta_0 \geq 0
\end{align*}
\]
Production possibility sets (available technology):

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

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

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

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

\[ \text{PPS} := \{(x, y) \mid y \text{ can be produced from } x\} \quad \text{(general PPS)} \]

Desirable properties for PPS

1. **convexity**: if \((x, y) \in \text{PPS}\) and \(\lambda \geq 0, \sum \lambda_k = 1\) then \((X\lambda, Y\lambda) \in \text{PPS}\)

2. **free (strong) disposability of inputs and outputs**:
   1. if \((x, y) \in \text{PPS}\) and \(x^+ := x + s^+\) with \(s^+ \geq 0\) then \((x^+, y) \in \text{PPS}\)
   2. if \((x, y) \in \text{PPS}\) and \(y^- := y - s^-\) with \(s^- \geq 0\) then \((x, y^-) \in \text{PPS}\)

3. **minimum intersection**: \(\text{PPS}\) is the intersection of all sets \(\overline{\text{PPS}}\) satisfying properties 1 and 2, subject to \((x, y) \in \overline{\text{PPS}}\).
Production possibility sets (available technology):

\[
PPS_L := \{ (x, y) \mid x = X\lambda, y = Y\lambda, \lambda \geq 0 \}
\]

\[
PPS_C := \{ (x, y) \mid x = X\lambda, y = Y\lambda, \sum \lambda_k = 1, \lambda \geq 0 \}
\]

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

\[
PPS_I := \{ (x, y) \mid x = X\lambda, y = Y\lambda, \sum \lambda_k = 1, \lambda \in \{0, 1\}^K \}
\]

\[
PPS := \{ (x, y) \mid y \text{ can be produced from } x \} \quad \text{(general PPS)}
\]

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 products (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

![Histogram of DEA and RDM results](attachment:image.png)
- 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
Further Step – Dynamic Behaviour
Malmquist Type Indexes

- **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) $\alpha$-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

\[
M_{t+1} := \sqrt{\frac{D_t(x_{0}^{t+1}, y_{0}^{t+1})}{D_t(x_{0}^{t}, y_{0}^{t})}} \cdot \frac{D_{t+1}(x_{0}^{t+1}, y_{0}^{t+1})}{D_{t+1}(x_{t}^{t}, y_{t}^{t})}
\]

two components of the Malmquist index:

\[
M_{t+1} := EC_{t+1} \cdot TC_{t+1}
\]

efficiency change \( EC_{t+1} := \frac{D_t(x_{0}^{t+1}, y_{0}^{t+1})}{D_t(x_{0}^{t}, y_{0}^{t})} \)

technology change \( TC_{t+1} := \sqrt{\frac{D_t(x_0^{t+1}, y_0^{t+1})}{D_{t+1}(x_{0}^{t+1}, y_{0}^{t+1})}} \cdot \frac{D_t(x_{t}^{t}, y_{t}^{t})}{D_{t+1}(x_{t}^{t}, y_{t}^{t})} \)


Index numbers and indifference surfaces

Farrel, M. J. (1957)
The Measurement of Productive Efficiency

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

Some models for estimating technical and scale inefficiencies in data envelopment analysis

DEA models for evaluating efficiency dominance
Research Report of the University of Texas, Center for Cybernetic Studies, Austin

Benefit and distance functions.
*Journal of Economic Theory, 70*(2), 407–419.
The economic theory of index numbers and the measurement of input, output, and productivity
*Econometrica*, 50(6), 1393–1414

*Journal of Productivity Analysis*, 3, 85–101

Productivity and undesirable outputs: a directional distance function approach
*Journal of Environmental Management*, 51(3), 229-240

A global Malmquist-Luenberger productivity index
*Journal of Productivity Analysis*, 34(3), 183-197

\(\alpha\)-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