Data Envelopment Analysis

Lucie Chytilová

VŠB-Technical University of Ostrava
Czech Republic

Letní škola aplikované informatiky Bedřichov
Bedřichov, 08.-09.9. 2017

Data Envelopment Analysis

Content

- Introduction
- Methodology
  - Classic DEA
  - Special DEA
- Example of Application
- Conclusion

Introduction

The aim

to use Data Envelopment Analysis in a problematic environment

- missing data
- dual-role variable
- risk variable

Overview of the Literature

- Charnes, Cooper and Rhodes (1978)
  - Measuring the efficiency of decision making units
- Banker, Charnes and Cooper (1984)
  - Extension of basic models
- Production process (variables)
  - production approach - Sherman and Gold (1985)
  - intermediation approach - Sealey and Lindley (1997)
- Missing variables - Smirlis et al. (2006)
- Dual-role variable - Cook et al. (2006)
- Risk component
  - exogenous treatment - Ataullah et al. (2004), Chang and Chiu (2006)
  - endogenous treatment - Chiu and Chen (2008), Girardone et al. (2004)

Data Envelopment Analysis

- Non-parametric technique based on linear programming for measuring the relative efficiency of a set of DMUs
- DEA models differs according to orientation or assumptions on return to scale
- An essential topic is how to choose the appropriate inputs and outputs
- The analysis of non-efficient units, projection of non-efficient units on the frontier etc.
- There exist many improvements of the classical DEA models - quasi fixed variables, fuzzy variables etc.

DEA

- the non-parametric approach for measuring the relative efficiency of the number of DMUs,
- let DMU_k for $k = 1, 2, ..., T$ and let input and output data for DMU_k be $\mathbf{X} = \{x_{ik}, i = 1, 2, ...R; k = 1, 2, ...T\}$ and $\mathbf{Y} = \{y_{jk}, j = 1, 2, ...S; k = 1, 2, ...T\}$, u_i for $i = 1, 2, ...R$ and v_j for $j = 1, 2, ...S$ be the weights of i^{th} input and j^{th} output, respectively,
- the relative efficiency score of DMU_k can be define as:

$$e_k = \frac{\sum_{j=1}^{S} v_j y_{jk}}{\sum_{i=1}^{R} u_i x_{ik}}, \text{ for } k = 1, 2, ..., T. \quad (1)$$

The multiplier input-oriented model - CCR

- Charmes et al. (1978),
- to measure the efficiency score of the under evaluation unit, DMU_Q where $Q \in \{1, ..., T\}$:

$$\begin{array}{lll} & \max\ e_Q = \sum_{j=1}^{S} v_j y_{jQ}, & \\ \text{s.t.} & \sum_{i=1}^{R} u_i x_{iQ} = 1, & \\ & \sum_{j=1}^{S} v_j y_{jk} - \sum_{i=1}^{R} u_i x_{ik} \leq 0, & k = 1, 2, ..., T, \\ & u_i \geq 0, & i = 1, 2, ..., R, \\ & v_j \geq 0, & j = 1, 2, ..., S. \end{array} \qquad (2)$$

- DMU_Q is CCR-efficient if and only if $e^* = 1$ and if there exists at least one optimal solution $(\overrightarrow{u^*}, \overrightarrow{v^*})$ with $\overrightarrow{u^*} > \overrightarrow{0}$ and $\overrightarrow{v^*} > \overrightarrow{0}$,
- inefficient units have a degree of relative efficiency less than one.

The multiplier input-oriented model - BCC

- Banker et al. (1984) extended the CCR model
- Convex envelope of data which leads to more efficient DMUs
- BCC model in dual multiplier form is mathematically as it follows:

$$
\begin{array}{lll}
 & \max\ e_Q = \sum_{j=1}^{S} v_j y_{jQ} - v_0, & \\
\text{s.t.} & \sum_{i=1}^{R} u_i x_{iQ} = 1, & \\
 & \sum_{j=1}^{S} v_j y_{jk} - \sum_{i=1}^{R} u_i x_{ik} - v_0 \leq 0, & k = 1, \ldots, T, \\
 & u_i \geq 0, & i = 1, \ldots, R, \\
 & v_j \geq 0, & j = 1, \ldots, S, \\
 & v_0 \in (-\infty, \infty), &
\end{array} \tag{3}
$$

where v_0 is the dual variable assigned to the convexity condition $\mathbf{e^T}\lambda = \mathbf{1}$ of envelopment form of BCC model

Data Envelopment Analysis with Missing Data

- Smirlis et al. (2006)
- $x_{ij} \in [x_{ij}^L, x_{ij}^U]$ and $y_{rj} \in [y_{rj}^L, y_{rj}^U]$
- $x_{ij} = x_{ij}^L + s_{ij}(x_{ij}^U - x_{ij}^L), \quad i = 1, ..., m; \; j = 1, ..., n$ with $0 \leq s_{ij} \leq 1$
  $y_{rj} = y_{rj}^L + t_{rj}(y_{rj}^U - y_{rj}^L) \quad r = 1, ..., s; \; j = 1, ..., n$ with $0 \leq t_{rj} \leq 1$
- model is mathematically as it follows:

$$\begin{array}{lll}
 & \max \; f_o(\mu, \mu_o) = \sum_{r=1}^{s} \mu_r \left(y_{rjo}^L + t_{rjo}(y_{rjo}^U - y_{rjo}^L)\right) - \mu_o, & \\
\text{s.t.} & \sum_{i=1}^{m} v_i \left(x_{ijo}^L + s_{ijo}(x_{ijo}^U - x_{ijo}^L)\right) = 1, & \\
 & \sum_{r=1}^{s} \mu_r \left(y_{rj}^L + t_{rj}(y_{rj}^U - y_{rj}^L)\right) - \sum_{i=1}^{m} v_i \left(x_{ij}^L + s_{ij}(x_{ij}^U - x_{ij}^L)\right) & \\
 & -\mu_o \leq 0, & j = 1, ..., n, \\
 & v_i \geq 0, 0 \leq s_{ij} \leq 1 & i = 1, ..., m, \\
 & \mu_j \geq 0, 0 \leq t_{rj} \leq 1 & r = 1, ..., s, \\
 & \mu_o \text{ free in sign.} &
\end{array} \tag{4}$$

Data Envelopment Analysis with Missing Data

- $q_{ij} = v_i s_{ij}$ and $p_{rj} = \mu_r t_{rj}$
- transformed model is mathematically as it follows:

$$
\begin{array}{lll}
 & \max\ f_o(\mu, \mu_o) = \sum_{r=1}^{s} \left(\mu_r y_{rjo}^L + p_{rjo}(y_{rjo}^U - y_{rjo}^L)\right) - \mu_o, & \\
\text{s.t.} & \sum_{i=1}^{m} \left(v_i x_{ijo}^L + q_{ijo}(x_{ijo}^U - x_{ijo}^L)\right) = 1, & \\
 & \sum_{r=1}^{s} \mu_r y_{rj}^L + \sum_{r=1}^{s} p_{rj}(y_{rj}^U - y_{rj}^L) - & \\
 & \sum_{i=1}^{m} v_i x_{ij}^L - \sum_{i=1}^{m} q_{ij}(x_{ij}^U - x_{ij}^L) - \mu_o \leq 0, & j = 1, ..., n, \\
 & q_{ij} - v_i \leq 0, & i = 1, ..., m, \\
 & p_{rj} - \mu_r \leq 0, & r = 1, ..., s, \\
 & u_i, \mu_r \geq 0, & \forall i, r, \\
 & q_{ij}, p_{rj}, \geq 0, & \forall i, r, j, \\
 & \mu_o \text{ free in sign.} &
\end{array}
\quad (5)
$$

Data Envelopment Analysis with Dual-Role Variable

- Cook et al. (2006)
- Linear programming form of CCR I-DEA dual-role variable problem:

$$
\begin{array}{lll}
 & \max\ e_Q = \sum_{j=1}^{S} v_j y_{jQ} + \gamma w_Q - \beta w_Q, & \\
\text{s.t.} & \sum_{i=1}^{R} u_i x_{iQ} = 1, & \\
 & \sum_{j=1}^{S} v_j y_{jk} + \gamma w_k - \beta w_k - \sum_{i=1}^{R} u_i x_{ik} \leq 0, & k = 1, 2, ..., T, \\
 & u_i \geq 0, & i = 1, 2, ..., R, \\
 & v_j \geq 0, & j = 1, 2, ..., S, \\
 & \gamma \geq 0, & \\
 & \beta \geq 0, &
\end{array}
\tag{6}
$$

where w_Q can serve as input, output or it can be in equilibrium and γ and β are weights for the dual-role variable

- if $\gamma^* - \beta^* < 0$ then the dual-role variable is "input variable"
- if $\gamma^* - \beta^* > 0$ then the dual-role variable is "output variable"
- if $\gamma^* - \beta^* = 0$, then the dual-role variable is at an equilibrium

New model I.

- where the dual-role variable is not precisely known

$$
\begin{array}{lll}
\max & f_o(\mu) = \sum_{r=1}^{s} \left(\mu_r y_{rjo}^L + p_{rjo}(y_{rjo}^U - y_{rjo}^L)\right) + & \\
& \gamma w_o^L + c_{jo}(w_o^U - w_o^L) - \beta w_o^L - d_{jo}(w_o^U - w_o^L), & \\
\text{s.t.} & \sum_{i=1}^{m} \left(v_i x_{ijo}^L + q_{ijo}(x_{ijo}^U - x_{ijo}^L)\right) = 1, & \\
& \sum_{r=1}^{s} \left(\mu_r y_{rj}^L + p_{rj}(y_{rj}^U - y_{rj}^L)\right) - \sum_{i=1}^{m} \left(v_i x_{ij}^L + q_{ij}(x_{ij}^U - x_{ij}^L)\right) & \\
& +\gamma w_j^L + c_j(w_j^U - w_j^L) - \beta w_j^L - d_j(w_j^U - w_j^L) \leq 0, & j = 1, ..., n, \\
& q_{ij} - v_i \leq 0, & i = 1, ..., m, \\
& p_{rj} - \mu_r \leq 0, & r = 1, ..., s, \\
& c_j - \gamma_j \leq 0, & \forall j \\
& b_j - \beta_j \leq 0, & \forall j \\
& u_i, \mu_r, \gamma, \beta \geq 0, & \forall i, r, \\
& q_{ij}, p_{rj}, c_j, d_j \geq 0, & \forall i, r, j
\end{array}
\quad (7)
$$

New model II.

- where the dual-role variable is precisely known

$$
\begin{array}{lll}
\max & f_o(\mu, \mu_o) = \sum_{r=1}^{s} \left(\mu_r y_{rjo}^L + p_{rjo}(y_{rjo}^U - y_{rjo}^L)\right) & \\
 & +\gamma w_o - \beta w_o, & \\
\text{s.t.} & \sum_{i=1}^{m} \left(v_i x_{ijo}^L + q_{ijo}(x_{ijo}^U - x_{ijo}^L)\right) = 1, & \\
 & \sum_{r=1}^{s} \mu_r y_{rj}^L + \sum_{r=1}^{s} p_{rj}(y_{rj}^U - y_{rj}^L) & \\
 & -\sum_{i=1}^{m} v_i x_{ij}^L - \sum_{i=1}^{m} q_{ij}(x_{ij}^U - x_{ij}^L) + \gamma w_o - \beta w_o \leq 0, & j = 1, ..., n, \\
 & q_{ij} - v_i \leq 0, & i = 1, ..., m, \\
 & p_{rj} - \mu_r \leq 0, & r = 1, ..., s, \\
 & u_i, \mu_r, \gamma, \beta \geq 0, & \forall i, r, \\
 & q_{ij}, p_{rj}, \geq 0, & \forall i, r, j
\end{array} \qquad (8)
$$

Input and Output Variables

- Visegrad Group (Czech Republic, Hungary, Poland, Slovakia)
- 35 banks (11, 7, 11, 6)
- Year 2015
- Intermediation approach
- Bankscope and Eurostat

Table 1: Description of Variables

Variables	Description in the balance sheet	Unit of measurement
Input Variables		
Physical Capital (x_1 - FA)	Fixed Assets	Euro
Labour (x_2 - LAB)	Number of Employees	Number
Loanable Funds (x_3 - LF)	Deposits + Short Term Funding	Euro
Output Variables		
Advances (y_1 - ADL)	Loans + Advances to Banks	Euro
Investments (y_2 - INV)	Other Securities	Euro
Non-Interest Income (y_3 - NII)	Non-Earning Assets	Euro

Conclusion

- the results give different results, but all are close to each other
- so far it looks that the model improves the results, but does not distort them

Thank you for your attention!