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

# Chance Constrained Data Envelopment Analysis

## The Productive Efficiency of Units with Stochastic Outputs

Michal Houda
Department of Applied Mathematics and Informatics

ROBUST 2016, September 11–16, 2016

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

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

## Definition 1

$\text{DMU}_1$ dominates $\text{DMU}_2$ 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

$\text{DMU}_0$ 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): $PPS_I = \{(x_k, y_k)\}_{k=1}^{K}$
Dominance wrt. $PPS_I$: additive model with integer constraints

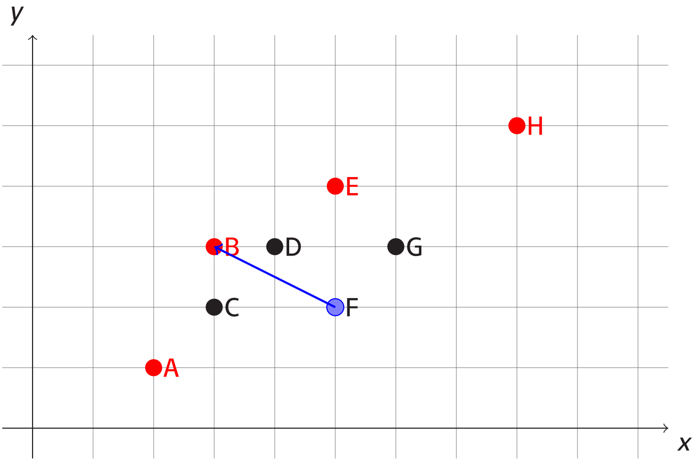
$$\max\Big(\sum_j s_j^+ + \sum_i s_i^-\Big) \text{ subject to}$$

$$\sum_k x_{ik}\lambda_k + s_i^- = x_{i0} \qquad \forall i \quad \text{(inputs)}$$

$$\sum_k y_{ik}\lambda_k - s_i^+ = y_{j0} \qquad \forall j \quad \text{(outputs)} \tag{1}$$

$$\sum_k \lambda_k = 1,\ \lambda_k \in \{0,1\}^K,\ s_i^-, s_j^+ \geq 0$$

- $s^-$ … slack for $X\lambda \leq x_0$ // $s^+$ … slack (surplus) for $Y\lambda \geq y_0$
- $\text{DMU}_0$ is efficient wrt. $PPS_I$ if no slack is greater than 0 (i. e., both inequalities are active) in optimal solution


y
x
H
E
B
D
G
C
F
A

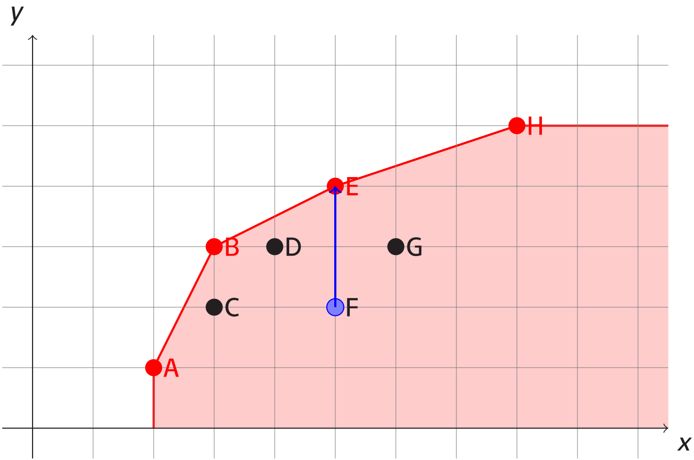
**Continuous (convex) PPS** (Banker, Cooper, Charnes, 1984):

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

Dominance wrt. $\mathrm{PPS}_C$: BCC (output oriented) model

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

$\epsilon$ … non-Archimedean infinitesimal



y
x
A
B
C
D
E
F
G
H

**Dual problem:**

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

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

**BCC (output oriented) DEA problem of fractional programming:**

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

## Definition 3 (DEA Efficiency)

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

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

**Remark**

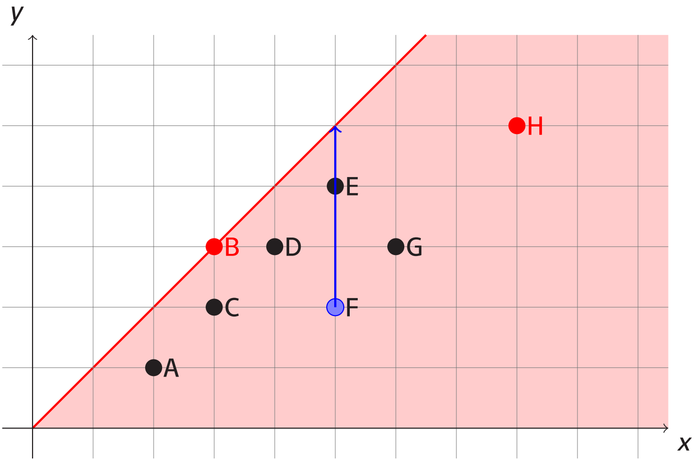
- weak DEA efficiency: $\phi^* = 1$ but some of $s_i^{-*}, s_j^{+*}$ are not zero (efficient points which are not extreme points of PPS)
- two-stage solution procedure:
  1. solve the BCC-O problem with $\epsilon = 0$ to obtain $\phi^*$
  2. solve the problem max $\sum_j s_j^+ + \sum_i s_i^-$ subject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

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

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

Dominance wrt. $\mathrm{PPS}_L$: CCR (output oriented) model

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



y
x
A
B
C
D
E
F
G
H

**Dual problem:**

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

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

**CCR (output oriented) DEA problem of fractional programming:**

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

## Introducing Randomness

- $X$, $Y$ are random matrices
- PPS random production possibility sets
- $\alpha \in [0; 1)$ ... tolerance (risk) level (sufficiently small)

Land, Lovell, Thore (1993), Cooper, Huang, Li (1996), Cooper, Huang et al (1998), Cooper, Deng, Huang, Li (2002), Cooper, Huang, Li (2004)

### Definition 4

DMU$_0$ is *not* stochastically dominated in its efficiency wrt. PPS$_I$ if $\forall \lambda \in \{0, 1\}^K$ with $\sum \lambda_k = 1$ we have

$$\mathbb{P}\{X\lambda \leq x_{\cdot 0},\ Y\lambda \geq y_{\cdot 0}\} \leq \alpha$$

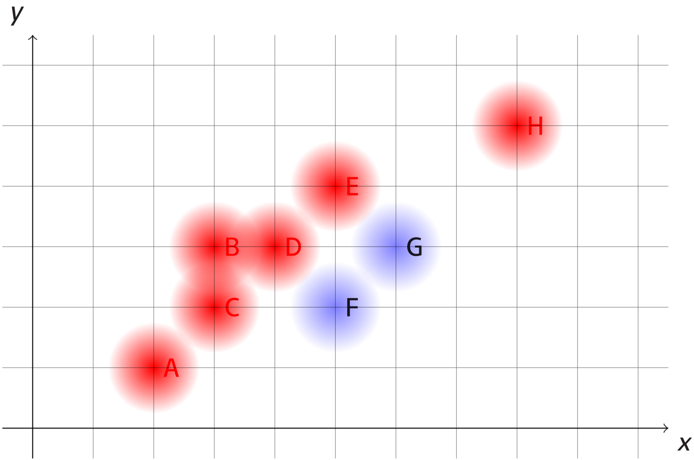
- if $X$, $Y$ are continuous we don't need "at least one strict" notion

**Stochastic extension of the additive 0-1 model**

$$\beta^* := \max \mathbb{P}\left\{X\lambda \leq x_{\cdot 0},\ Y\lambda \geq y_{\cdot 0}\right\}$$

$\text{DMU}_0$ is stochastically dominated $\Leftrightarrow \beta^* > \alpha$.

Recall $\text{PPS}_l = \left\{(x_k, y_k)\right\}_{k=1}^{K}$ (random discrete production set).



y
x
A
B
C
D
E
F
G
H

## Definition 5

1. $(x^*, y^*) \in \mathrm{PPS}_C$ is $\alpha$-stochastically efficient wrt. $\mathrm{PPS}_C$ if $\forall \lambda \geq 0$ with $\sum_k \lambda_k = 1$ we have

$$\mathbb{P}\left\{X\lambda \leq x^*,\ Y\lambda \geq y^*\right\} \leq \alpha$$

2. the set of all $\alpha$-stochastically efficient points is called $\alpha$-stochastically efficient frontier of $\mathrm{PPS}_C$
3. $\mathrm{DMU}_0$ is $\alpha$-stochastically efficient wrt. $\mathrm{PPS}_C$ if $\forall \lambda \in \{0,1\}^K$ with $\sum \lambda_k = 1$ we have

$$\mathbb{P}\left\{X\lambda \leq x_{\cdot 0},\ Y\lambda \geq y_{\cdot 0}\right\} \leq \alpha$$

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

## Testing efficiency of $\mathrm{DMU}_0$

### Proposition 6 (sufficient condition)

*If*

$$\mathbb{P}\{\mathbf{1}^T(X\lambda - x_0) < \mathbf{1}^T(Y\lambda - y_{\cdot 0})\} \leq \alpha$$

*then* $\mathrm{DMU}_0$ *is* $\alpha$*-stochastically efficient.*

### Proposition 7 (necessary condition)

*If* $\mathrm{DMU}_0$ *is* $\alpha$*-stochastically efficient then* $\forall\lambda \geq 0$ *with* $\sum \lambda_k = 1$ *and*

$$\mathbb{P}\{x_{i\cdot}\lambda < x_{i0}\} \geq 1 - \epsilon \qquad \forall i \quad \textit{(inputs)}$$
$$\mathbb{P}\{y_{j\cdot}\lambda > y_{i0}\} \geq 1 - \epsilon \qquad \forall j \quad \textit{(outputs)}$$

*then*

$$\mathbb{P}\{\mathbf{1}^T(X\lambda - x_0) < \mathbf{1}^T(Y\lambda - y_{\cdot 0})\} \leq \alpha$$

## Almost 100% Confidence Chance-Constrained Problem

$$\begin{aligned}\beta^* := \max \mathbb{P}\{\mathbf{1}^T(X\lambda - x_0) + \mathbf{1}^T(y_{\cdot 0} - Y\lambda) < 0\} \quad & \text{subject to} \\ \mathbb{P}\{x_{i\cdot}\lambda < x_{i0}\} \geq 1 - \epsilon \qquad & \forall i \quad \text{(inputs)} \\ \mathbb{P}\{y_{j\cdot}\lambda > y_{i0}\} \geq 1 - \epsilon \qquad & \forall j \quad \text{(outputs)} \\ \lambda \geq 0, \sum \lambda_k = 1 \qquad &\end{aligned}$$

### Proposition 8

1. *If* $\text{DMU}_0$ *is* $\alpha$*-stochastically efficient then* $\beta^* \leq \alpha$.
2. *If* $\beta^* > \alpha$ *then* $\mathit{DMU}_0$ *is not* $\alpha$*-stochastically efficient.*

**Example: normally distributed outputs**

- let $Y$ follow multivariate normal distribution;
- denote $\bar{X} := \mathbb{E}X$, $\bar{Y} := \mathbb{E}Y$, and $\sigma_j^2(\lambda) := \mathrm{var}(y_{j\cdot}^T\lambda - y_{j0})$

$$\beta^* := \min \mathbf{1}^T(X\lambda - x_0) + \mathbf{1}^T(\bar{y}_{\cdot 0} - \bar{Y}\lambda) + \sigma_0(\lambda)\Phi^{(-1)}(\alpha) \quad \text{subject to}$$
$$x_{i\cdot}\lambda \le x_{i0} \qquad \forall i \quad \text{(inputs)}$$
$$\bar{y}_{j\cdot}\lambda + \sigma_k(\lambda)\Phi^{(-1)}(\epsilon) \ge \bar{y}_{j\cdot} \qquad \forall j \quad \text{(outputs)}$$
$$\lambda \ge 0, \sum \lambda_k = 1$$

## Proposition 9

1. *If* $\mathrm{DMU}_0$ *is $\alpha$-stochastically efficient then* $\beta^* \ge 0$.
2. *If $\beta^* < 0$ then $DMU_0$ is not $\alpha$-stochastically efficient.*

Cooper, Deng, Huang, Li (2002, 2003): $\mathbb{E}$-model for BCC

$$\max \phi + \epsilon\Big(\sum_j s_j^+ + \sum_i s_i^-\Big) \quad \text{subject to}$$

$$\begin{aligned} \mathbb{P}\{x_{i\cdot}\lambda + s_i^- \le x_{i0}\} &\ge 1-\alpha && \forall i \quad \text{(inputs)} \\ \mathbb{P}\{y_{j\cdot}\lambda - s_j^+ \ge \phi y_{i0}\} &\ge 1-\alpha && \forall j \quad \text{(outputs)} \\ \lambda \ge 0, \sum \lambda_k &= 1 \end{aligned} \tag{8}$$

## Definition 10

$\text{DMU}_0$ is $\alpha$-stochastically DEA efficient if

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

- let $Y$ follow multivariate normal distribution;
- let $\bar{Y} := \mathbb{E}Y$, and $\sigma_j^2(\phi, \lambda) := \mathrm{var}(y_{j\cdot}^T \lambda - \phi y_{j0})$

Then the (nonlinear) DEA problem to solve is

$$\max \phi + \epsilon\Big(\sum_j s_j^+ + \sum_i s_i^-\Big) \quad \text{subject to}$$

$$\begin{aligned} x_{i\cdot}\lambda + s_i^- &= x_{i0} && \forall i \quad \text{(inputs)} \\ \bar{y}_{j\cdot}^T\lambda + \Phi^{(-1)}(\alpha)\sigma_j(\phi, \lambda) - s_j^+ &= \phi \bar{y}_{j0} && \forall j \quad \text{(outputs)} \\ \sum \lambda_k = 1,\ \lambda_k \geq 0,\ s_i^-, s_j^+ &\geq 0 \end{aligned} \tag{9}$$

(DEA problem with adjusted outputs).

We can use second-order cone programming approximation scheme to find the upper and lower bounds for the problem (c.f. Cheng, Houda, Lisser (2014)).

# References

Cooper, W. W., Seiford, L. M. and Zhu, J., eds. (2004).
*Handbook on Data Envelopment Analysis*.
Kluwer, New York.

Cooper, W. W., Deng, H., Huang, Z. and Li, S. X. (2002)
Chance constrained programming approaches to technical efficiencies and inefficiencies in stochastic data envelopment analysis.
*Journal of the Operational Research Society*, **53**(12), 1347–1356.

Cheng, J., Houda, M. and Lisser, A. (2014).
Second-order cone programming approach for elliptically distributed joint probabilistic constraints with dependent rows.
*Optimization Online*, Paper No. 4363, May 2014.