Data Envelopment Analysis

Gianfranco Atzeni

Laboratorio di valutazione delle decisioni di investimento

Introduction to DEA

DEA is a method for measuring efficiency of Decision-Making Units (DMU) using linear programming techniques to envelop observed input–output vectors as tightly as possible.

DEA allows multiple inputs–outputs to be considered at the same time without any assumption on data distribution:

- in both cases efficiency is measured in terms of a proportional change in inputs or outputs.

A DEA model can be subdivided into an:

- **input-oriented** model, which minimizes inputs while satisfying at least the given output levels;
- **output-oriented** model, which maximizes outputs without requiring more of any observed input values.

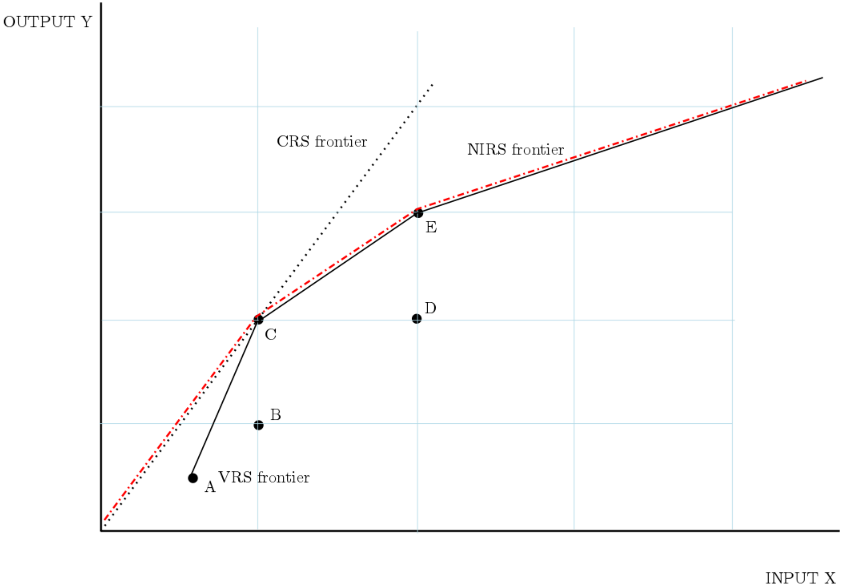
Introduction to DEA

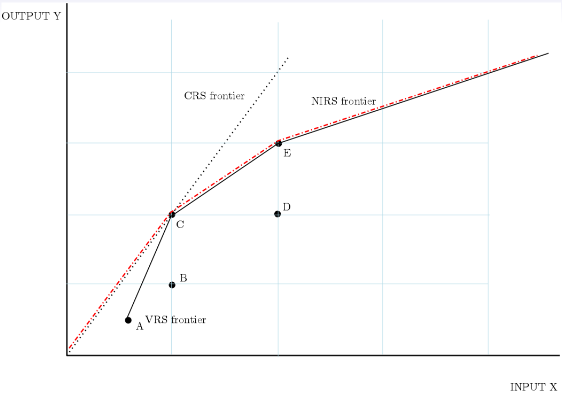
DEA models can also be subdivided in terms of returns to scale.

Charnes, Cooper, and Rhodes (1978) originally proposed the efficiency measurement of the DMUs for **constant returns to scale** (**CRS**), where all DMUs are operating at their optimal scale.

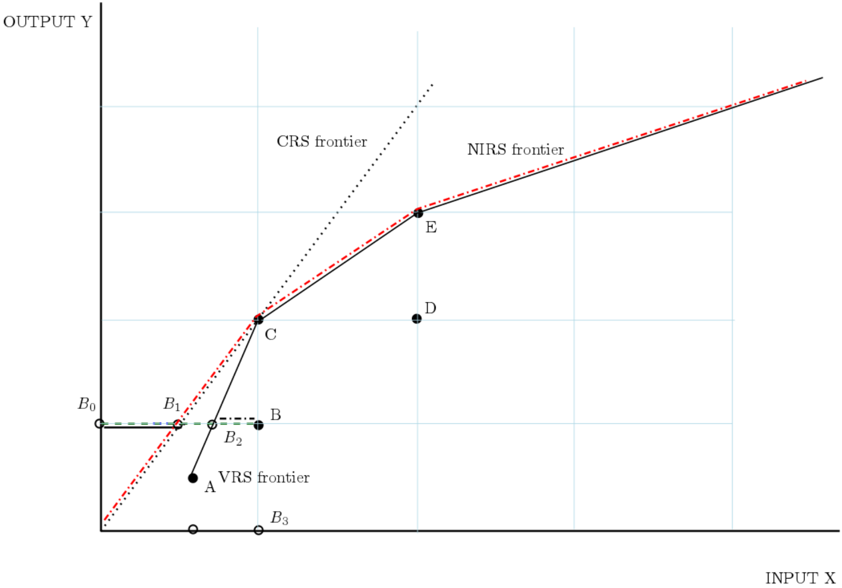
Later Banker, Charnes, and Cooper (1984) introduced the **variable returns to scale** (VRS) efficiency measurement model, allowing the breakdown of efficiency into **technical** and **scale efficiencies** in DEA.

It is possible to illustrate this by drawing the frontiers determined by economies of scale with one input and one output.


OUTPUT Y
CRS frontier
NIRS frontier
E
D
C
B
A
VRS frontier
INPUT X



- If CRS are assumed then the only efficient DMU is C.
- A, C, E are efficient only if VRS are assumed.
- Where the nonincreasing returns to scale (dash dotted red line) and VRS frontiers are equal, we have decreasing returns to scale for those DMUs on the efficient frontier (such as E)


OUTPUT Y
CRS frontier
NIRS frontier
E
D
C
B_0
B_1
B
B_2
A
VRS frontier
B_3
INPUT X

Efficiency

For input oriented DEA with CRS efficiency of DMU B is defined as:

$$\theta_{B,INPUT,CRS} = \frac{\overline{B_0B_1}}{\overline{B_0B}} \quad (1)$$

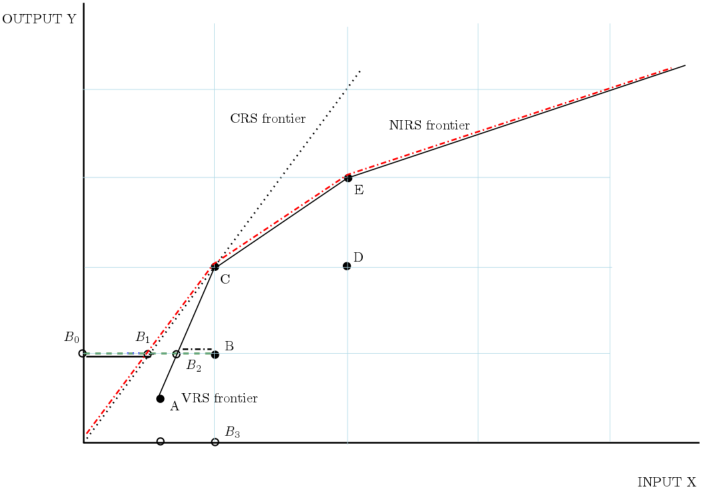
It represents that one can obtain the same output by reducing the input by the ratio of $1 - \theta_{B,INPUT,CRS}$

Accordingly, the input-oriented efficiency relative to the VRS frontier is defined as:

$$\theta_{B,INPUT,VRS} = \frac{\overline{B_0B_2}}{\overline{B_0B}} \quad (2)$$

Efficiency

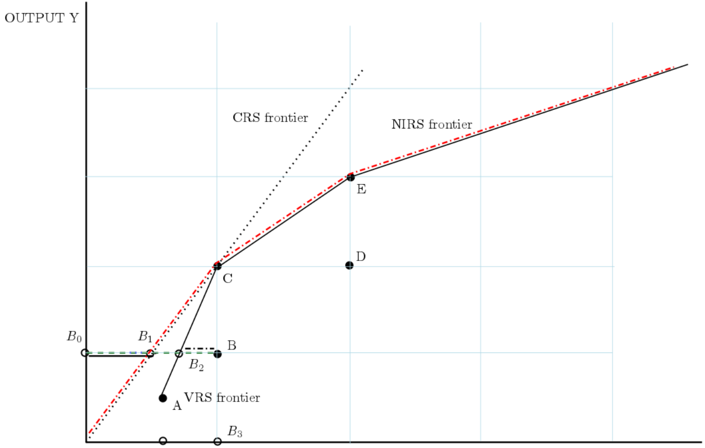
All efficiency measures of DMU C are the same because the frontiers meet at point C.



Efficiency output oriented

The efficiency for the output-oriented CRS DEA model is defined as:

$$\theta_{B,INPUT,VRS} = \frac{\overline{B_3B}}{\overline{B_3C}} \quad (3)$$



Decomposition of technical inefficiency

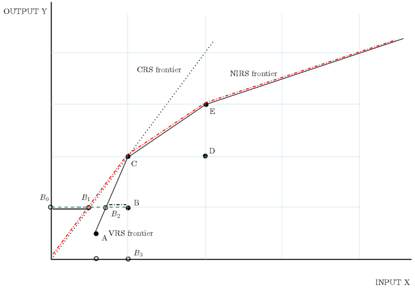
It is possible

to decompose the CRS technical

inefficiency into scale efficiency

and ”pure” technical efficiency.

For point B:



$\overline{B_2 B}$ contributes

to the **technical efficiency**

regarding the VRS model.

$\overline{B_1 B}$ contributes to the **technical efficiency** regarding the CRS model.

$\overline{B_1 B_2}$ contributes to the **scale efficiency**.

Efficiency, slack, peers

We measure efficiency as the relative distance to frontier:

- firms that are technically inefficient operate at points in the interior of the frontier;
- while those that are technically efficient operate somewhere along the technology defined by the frontier.

The DMU is called efficient when the DEA score (θ) equals 1 and all slacks are 0:

- If only the first condition is satisfied we have **”weak” efficiency**;
- If both conditions are satisfied we have **”strong” efficiency** (Pareto–Koopman efficiency).

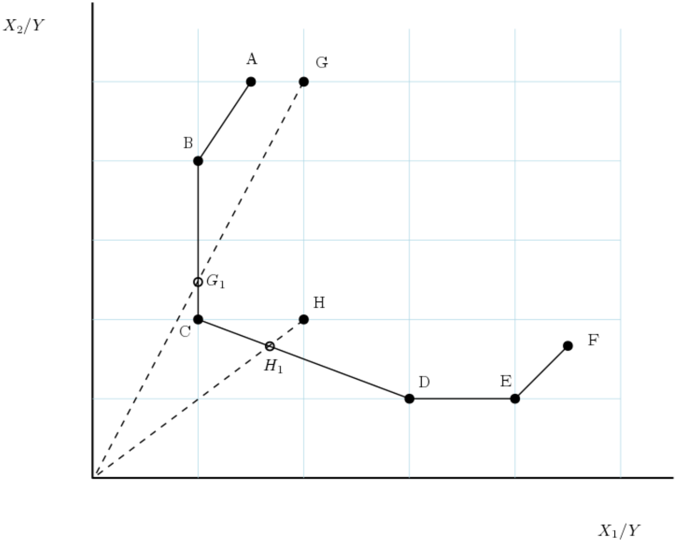
Slacks

Consider DMU G and H in the figure, where is depicted a production with two inputs and one output. Their technical efficiencies are defined as $\overline{OG_1}/\overline{OG}$ and $\overline{OH_1}/\overline{OH}$, respectively.

For DMU G, point G_1 is the efficient point, however input X_2 can be further reduced. We say that DMU G has **input slack** equal to $\overline{CG_1}$. That means, the reduction of input necessary to move from ”weak efficiency” to a point of ”strong efficiency”.

Reference or peer is a point that an inefficient DMU, such as point G, targets to move from the efficient point G_1, to the Pareto–Koopman efficient point C.

Slacks, Reference or Peers



Output maximizing programme

With n DMUs (labelled by j), m inputs (labelled by i) and p outputs (labelled by r), the maximizing programme is:

$$max \;\; d_0 = \frac{\sum_{r=1}^{p} u_r y_{r0}}{\sum_{i=1}^{m} v_i xij}$$

$$s.t.$$

$$\frac{\sum_{r=1}^{p} u_r y_{rj}}{\sum_{i=1}^{m} v_i xij} \leq 1 \qquad j = 1...n$$

where: y_{r0} is quantity of output r for DMU 0; u_r is the weight attached to output r, $u_r > 0$, with $r = 1, ..., p$

x_{i0} is the quantity of input i for DMU 0; v_i is the weight attached to input i, $v_i > 0$, with $i = 1, ..., m$

Output maximizing programme

The weights are specific to each unit so that $0 \leq d_0 \leq 1$ and a value of unity implies complete technical efficiency relative to the sample of units under scrutiny.

The weights are not known a priori, they are calculated from the efficiency frontier by comparing a particular DMU with other ones producing similar outputs and using similar inputs (DMU's peers).

DEA computes all possible sets of weights which satisfy all constraints and chooses those which give the most favorable view of the DMU, that is the highest efficiency score. The numerator or the denominator of the efficiency ratio are constrained to be equal to one, so that $0 \leq d_0 \leq 1$

Input minimizing programme

The problem of maximizing weighted output with weighted input equal to one becomes one of minimizing weighted input with weighted output equal to one. The dual minimizing programme is:

$$min \;\; \theta = Z$$

$$s.t$$

$$\sum_{j=1}^{n} x_{ij}\lambda_j \leq x_{i0}Z \qquad j = 1...n$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{r0} \qquad j = 1...n$$

where: $\lambda_j \geq 0$ and $j = 1, ..., n$. λ_j are weights on units sought to form a composite unit to outperform j_0.

Input minimizing programme

The model is solved giving each DMU in the sample an efficiency score.

The model computes the factor Z needed to reduce the input of DMU 0 to a frontier formed by its peers (or convex combinations of them) which produce no less output than DMU 0 and use a fraction Z of input of DMU 0.

The DMU will be efficient if $Z = 1$. In other words a composite unit cannot be constructed which outperforms it.

If $Z < 1$, the DMU will be inefficient. The composite unit provides targets for the inefficient unit and Z represents the maximum inputs a DMU should be using to attain at least its current output.

Variable return to scale

DEA can be carried out with either the constant or variable returns to scale assumption (CRS or VRS). The model above is consistent with the CRS production frontier.

To calculate the VRS frontier a further constraint is included:

$$\sum_{j=1}^{n} \lambda_j = 1 \tag{4}$$

The VRS approach produces technical efficiency scores which are greater than or equal to those obtained using CRS and is therefore the more flexible assumption of the underlying production technology.

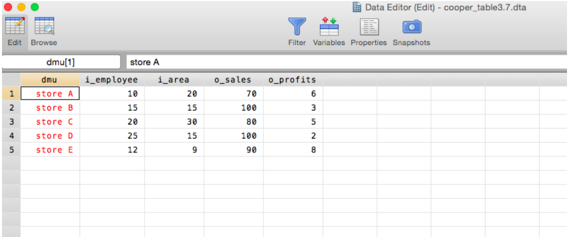
CRS input-oriented two-stage DEA with STATA

We learn how to use the STATA command **dea**. The default command gives the CRS input-oriented two stage DEA.

Two stage means that we consider the slacks of the DMUs. If we disregard the slack and calculate it residually, the model becomes the single-stage DEA model.

The dea command requires an initial dataset that contains the input and output variables for observed DMU. Variable names must be identified by ivars for input variables and by *ovars* for output variables so that the **dea** command can identify and handle the multiple input–output dataset.

For this example we use the dataset `cooper_table3.7.dta`. The data are organized as follows



	dmu	i_employee	i_area	o_sales	o_profits
1	store A	10	20	70	6
2	store B	15	15	100	3
3	store C	20	30	80	5
4	store D	25	15	100	2
5	store E	12	9	90	8

We have five stores that use two inputs: `i_employees` (number of employees) and `i_area` (the area of floor)
to produce two outputs: `o_sales` (the volume of sales) and `o_profits` (the volume of profits).

dea i_employee i_area = o_sales o_profits

```
      name:  dealog
       log:  /Users/gianfranco/Dropbox/LAVED/DEA/dea.log
  log type:  text
 opened on:   6 Feb 2017, 15:54:39

options: RTS(CRS) ORT(IN) STAGE(2)
CRS-INPUT Oriented DEA Efficiency Results:
                                              ref:        ref:        ref:        ref:        ref:     islack:
                    rank       theta     store_A     store_B     store_C     store_D     store_E  i_employee
dmu:store_A            2     .933333           .           .           .           .     .777778           .
dmu:store_B            3     .888889           .           .           .           .     1.11111           .
dmu:store_C            5     .533333           .           .           .           .     .888889           .
dmu:store_D            4     .666667           .           .           .           .     1.11111     3.33333
dmu:store_E            1           1           .           .           .           .           1           .

                  islack:     oslack:     oslack:
                   i_area     o_sales   o_profits
dmu:store_A       11.6667    2.98e-07     .222222
dmu:store_B       3.33333    7.45e-07     5.88889
dmu:store_C             8    4.17e-06     2.11111
dmu:store_D             .    2.98e-06     6.88889
dmu:store_E             0           .           .
      name:  dealog
       log:  /Users/gianfranco/Dropbox/LAVED/DEA/dea.log
  log type:  text
 closed on:   6 Feb 2017, 15:54:39
```

Interpretation of results

Store E is the only efficient DMU and is the referent for all other stores.

Consider store A:

- the optimal solution of efficiency score (`theta`) is .933333
- reference weights (λ_A λ_B λ_C λ_D λ_E) are $(0, 0, 0, 0, 0.777778)$
- slack (`i_area, i_employee, o_sales, o_profits`) are $(11.6667, 0, 0, 0, 0.222222)$

Score. Store A has an efficient score of 93.33%, then all inputs (*employee* and *area*) could be reduced by 6.67%.

Slack. Store A has an input slack for *area* of 11.67. This means that 11.67 units of *area* could be reduced even after Store A has reduced all inputs by 6.67%.

VRS input-oriented DEA

For this example we use the dataset in `coelli table6.4.dta`. Data consist of five firms that use one input, `i_1`, to produce one output, `o_1`.

The command to **`dea i_x = o_q, rts(vrs) stage(1)`**

```
      name:  dealog
       log:  /Users/gianfranco/Dropbox/LAVED/DEA/dea.log
  log type:  text
 opened on:   6 Feb 2017, 19:03:36

options: RTS(VRS) ORT(IN) STAGE(1)
VRS-INPUT Oriented DEA Efficiency Results:
                                ref:     ref:     ref:     ref:     ref:  islack:  oslack:
            rank    theta          A        B        C        D        E      i_x      o_q
dmu:A          1        1          1        .        .        .        .        .        0
dmu:B          5     .625         .5        .       .5        .        .        .        .
dmu:C          1        1          0        .        1        .        .        .        .
dmu:D          4       .9          .        .       .5        .       .5        .        .
dmu:E          1        1          .        .        0        .        1        .        .


VRS Frontier(-1:drs, 0:crs, 1:irs)
          CRS_TE     VRS_TE    NIRS_TE      SCALE         RTS
dmu:A   0.500000   1.000000   0.500000   0.500000    1.000000
dmu:B   0.500000   0.625000   0.500000   0.800000    1.000000
dmu:C   1.000000   1.000000   1.000000   1.000000    0.000000
dmu:D   0.800000   0.900000   0.900000   0.888889   -1.000000
dmu:E   0.833333   1.000000   1.000000   0.833333   -1.000000


VRS Frontier:

     +----------------------------------------------------------+
     | dmu   o_q   i_x     CRS_TE     VRS_TE      SCALE     RTS |
     |----------------------------------------------------------|
  1. |   A     1     2   0.500000   1.000000   0.500000     irs |
  2. |   B     2     4   0.500000   0.625000   0.800000     irs |
  3. |   C     3     3   1.000000   1.000000   1.000000       - |
  4. |   D     4     5   0.800000   0.900000   0.888889     drs |
  5. |   E     5     6   0.833333   1.000000   0.833333     drs |
     +----------------------------------------------------------+
      name:  dealog
       log:  /Users/gianfranco/Dropbox/LAVED/DEA/dea.log
  log type:  text
 closed on:   6 Feb 2017, 19:03:36
```

Interpretation of results

The efficiency score (`theta`) of DMU B is 0.625, and DMUs A and C are the reference DMUs for DMU B.

The sum of the reference weights should equal 1 because `rts(vrs)` specifies that $\sum_{j=1}^{n} \lambda_j = 1$. The sum of the reference weights for DMU B equals 1 as:

$(\lambda_A, \lambda_B, \lambda_C, \lambda_D, \lambda_E)$=(0.5, 0, 0.5, 0, 0).

Stores A, C, and E are the efficient points that inefficient DMUs (B and D) can target to move in input-oriented DEA calculation.

Additional information with `rts(vrs)`: stores D and E are on the decreasing returns to scale (`drs`) portion of the VRS frontier. On the other hand, Stores A and B are on the increasing returns to scale (`irs`) portion of the VRS frontier.