EMS: Efficiency Measurement System User's Manual

Holger Scheel

Version 1.3
2000-08-15

Contents

1 Introduction

Efficiency Measurement System (EMS) is a software for Windows 9x/NT which computes Data Envelopment Analysis (DEA) efficiency measures. This manual is intended to be an introduction to the usage of the software. It is not an introduction to DEA which you can found e. g. in the following books:

- H. O. FRIED, C. A. K. LOVELL, and S. SCHMIDT (1993), *The measurement of productive efficiency: Techniques and applications*, Oxford University Press, New York

- R. FÄRE, S. GROSSKOPF, and C. A. K. LOVELL (1994), *Production Frontiers*, Cambridge University Press, Cambridge

- A. CHARNES, W. W. COOPER, A. Y. LEWIN, and L. M. SEIFORD (1994), *Data Envelopment Analysis: Theory, Methodology, and Application*, Kluwer Academic Publishers, Dordrecht

- W. W. COOPER, L. M. SEIFORD, and K. TONE (2000), *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*, Kluwer Academic Publishers, Norwell, Massachusetts

The latest news about EMS, downloads and bugfixes you'll find on the EMS homepage:

<div align="center">

http://www.wiso.uni-dortmund.de/lsfg/or/scheel/ems/

</div>

EMS uses the LP Solver DLL BPMPD 2.11 by Csaba Mészáros for the computation of the scores (Sources: http://www.netlib.org). It is an interior point solver.

If you have questions which are not answered in the following paragraphs or if you have suggestions for further developments send an email to

<div align="center">

H.Scheel@wiso.uni-dortmund.de

</div>

2 Preparing the input output data

The first and probably most difficult step in an efficiency evaluation is to decide which input and output data should be included. EMS accepts data in MS Excel or in text format. Additionally to "standard" inputs and outputs EMS can also handle "nondiscretionary" inputs and outputs (i. e., data which are not controlled by the DMUs). The next sections describe how the data files should be prepared for EMS. The size of your analysis is limited by the memory of your PC. I. e., there is theoretically no limitation of the number of DMUs, inputs and outputs in EMS. Although the code is not optimized for large scale data, we successfully solved problems with over 5000 DMUs and about 40 inputs and outputs. (Please let me know your experience with larger datasets.)

2.1   Using MS Excel files

EMS accepts Excel 97 (and older) files (`*.xls`). The input output data should be collected in one worksheet. Don't use formulas in this sheet, it should only contain the pure data and nothing else. EMS needs the following data format:

```
Data
{I}
{O}
```

- The name of the worksheet must be "`Data`".

- The first line contains the input/output names. First inputs, then outputs.

- Input names contain the string "`{I}`".

- Output names contain the string "`{O}`".

- The first column contains the DMU names.

Cf. the example file `EXAMPLE.XLS`

2.2   Using textfiles

For those who prefer another spreadsheet software than MS Excel, EMS accepts also plain textfiles (`*.txt`). For reading textfiles correctly EMS needs the file `schema.ini` which contains some formatting information. The following is necessary for using textfiles with EMS:

```
schema.ini
```

- Put `schema.ini` in the same directory where your textfiles are.

- Modify `schema.ini` by replacing "`[Yourfile.txt]`" by the name of your file.

The textfile which contains the input output data should then satisfy the following:

- Columns are separated by Tabs. **Make sure that exactly one Tab appears between two columns and that you don't have Tabs at other places in the file (e. g. at the end)**. You can check this in a texteditor by making the Tabs "visible".

- Input names contain the string "`{I}`".

- Output names contain the string "{O}".

- The first column contains the DMU names.

Cf. the example file `EXAMPLE.TXT`.

2.3   Non-discretionary inputs and outputs

EMS accepts non-discretionary data if in the data file the corresponding input name contains "{IN}" instead of "{I}", or the corresponding output name contains "{ON}" instead of "{O}".

> ```
> {IN}
> {ON}
> ```

When EMS computes an efficiency score (which is a distance to the efficient frontier) it doesn't alter the values of non-discretionary data. I. e., the distance will only be computed in the directions of the "normal" (descretionary) inputs and outputs while the non-discretionary are fixed.

Literature: EMS uses the idea of R. D. Banker and R. C. Morey (1986), Efficiency Analysis for Exogenously Fixed Inputs and Outputs, *Operations Research*, **34**, 513–521. See also for an overview: M. Staat (1999), Treating non-discretionary variables one way or the other: Implications for efficiency scores and their interpretation, In G. Westermann (ed.), *Data Envelopment Analysis in the Service Sector*, pp. 23–50, Gabler, Wiesbaden.

3   Preparing weights restrictions

You can specify weights restrictions of the form $W(p, q) \geq 0$, where p is the vector of input weights and q is the vector of output weights (or shadow prices). Hence, you can incorporate both "Cone Ratio" constraints and "Assurance Region" constraints.

Example. Suppose you have 3 inputs and 2 outputs and you want to have the restriction $p_1 \geq p_2$ then corresponding row in the weights restriction matrix W is $(1; -1; 0; 0; 0)$. If you have in addition bounds on the marginal rates of substitutions like $0.3 \leq \frac{q_1}{q_2} \leq 3$, then you transform them into two constraints $q_1 - 0.3q_2 \geq 0$ and $-q_1 + 3q_2 \geq 0$, yielding the rows $(0; 0; 0; 1; -0.3)$ and $(0; 0; 0; -1; 3)$ in the matrix W. Thus in this example one has

$$W = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -0.3 \\ 0 & 0 & 0 & -1 & 3 \end{pmatrix}$$

Like the input output data EMS accepts weights restriction data W in MS Excel and textfiles.

3.1   Using MS Excel files

EMS accepts Excel 97 (and older) files (`*.xls`). The weights restriction data should be collected in one worksheet. Don't use formulas in this sheet, it should only contain the pure data and nothing else. EMS needs the following data format:

`Weights`

- The name of the worksheet must be "`Weights`". (It *can* be contained in the same file as the "`Data`" sheet, but you may also choose another file.)

- The first row (the input/output names) should be *identical* to the corresponding `Data` sheet.

- The first column contains a name for each restriction.

Cf. the example file `EXAMPLE.XLS`

3.2   Using textfiles

If you have W in a textfile, you'll have to do the same like for the input output data: First put the file `schema.ini` in the directory of your textfiles and modify it, i. e., replace "`[Yourweightfile.txt]`" by the name of your file.

The textfile which contains W should then satisfy the following:

- Columns are separated by Tabs.

- The first row (the input/output names) should be *identical* to the corresponding input output data file.

- The first column contains a name for each restriction.

Cf. the example file `WEIGHTS.TXT`.

Literature: See for an overview: R. ALLEN, A. ATHANASSOPOULOS, R. DYSON, and E. THANASSOULIS (1997), Weights restrictions and value judgements in Data Envelopment Analysis: Evolution, development and future directions, *Annals of Operations Research*, **73**, 13–34.

4   Starting EMS and loading data

When you've prepared the data in Excel and/or textfiles as described above you can start EMS by clicking on it in the program folder.

> Menu `File`
> `Load data (Ctrl+O)`

Now you should connect EMS to the data: Your input output data can be loaded by pressing `Ctrl+O` (Menu `File` → `Load data`). If you select an appropriate filename then EMS tries to connect to this file. For large scale DEA evaluations with thousands of DMUs this connection may need a few seconds. The connection was successful

- if the filename appears in the statusbar (at the bottom of the EMS window) and

- the sand clock vanishes.

EMS does not display your data! If you want to edit your data, you should open the datafile in Excel or in your texteditor and edit it there. You should save the changes (it's not necessary to close the file) and then `Load data (Ctrl+O)` in EMS again. EMS always loads the file version from the harddisk.

> Menu `File`
> `Load Weight Restr (Ctrl+W)`

The file which contains the matrix W can be loaded by pressing `Ctrl+W` (Menu `File` → `Load Weight Restr`). When the file is successfully loaded its name is displayed in the statusbar (like the input output data file).

5   Running a DEA model

> `bpmpd.par`

Before running a DEA model, make sure that the file `bpmpd.par` is in the same folder as your data file!

5.1   Preparing the results format

> Menu `DEA`
> `Format (Ctrl+F)`

`Ctrl+F` (Menu `DEA` → `Format`) will display the Format dialog. Here, you may specify the number of decimals to display in the results table wich will be produced by EMS.

Moreover, you can decide whether

- the pure input and ouput weights (shadow prices) p_i and q_j should be displayed (Option `pure weights`) or

- the "virtual inputs and outputs", i.e. the weights multiplied by the input and output values $p_i \cdot x_{i0}$, $q_j \cdot y_{j0}$ (Option `virtual inputs/outputs`)

should be displayed in the results table.

5.2 Choosing a technology structure

`Ctrl`+M (Menu `DEA` $\rightarrow$ `Run model`) will display a dialog where you can specify the model you want to compute. Under `Models` you may choose between various technology structures:

```
Menu DEA
Run model (Ctrl+M)
Structure
Returns to Scale
```

- convex and nonconvex envelopment,

- constant, variable, nonincreasing or nondecreasing returns to scale.

5.3 Choosing an efficiency measure

An efficiency measure quantifies in one way or another a "distance" to the efficient frontier of the technology. EMS allows computation of various distances in input-, output- and non-oriented versions.

```
Orientation
```

An input oriented measure quantifies the input reduction which is necessary to become efficient holding the outputs constant. Symmetrically, an output oriented measure quantifies the necessary output expansion holding the inputs constant. A non-oriented measure quantifies necessary improvements when both inputs and outputs can be improved simultaneously.

It seems that in applications the choose of a certain measure mostly depends on three criteria:

- The "primal" interpretation, i.e. the meaning of the efficiency score with respect to input and output *quantities*,

- the "dual" interpretation, i. e. the meaning of the efficiency score with respect to input and output *prices*,

- the axiomatic properties of the efficiency measure (e. g. monotonicity, units invariance, indication of efficiency, continuity).

Most of the measures are similar with respect to these criteria, whence in this manual only essential differences are mentioned roughly[1] when the measures are defined below. $\mathcal{T}$ denotes the technology and (X^k, Y^k) denotes the input output data of the DMU under evaluation.

Distance

Radial: This measure (a.k.a. Debreu-Farrell-measure, or "radial part" of the CCR/BCC measure) indicates the necessary improvements when all relevant factors are improved by the same factor equiproportionally. Its oriented versions have nice price interpretations (cost reduction/revenue increase), but it doesn't indicate Koopmans efficiency.

$$\text{non-oriented:} \quad \max\{\theta \mid ((1-\theta)X^k, (1+\theta)Y^k) \in \mathcal{T}\}$$
$$\text{input:} \quad \min\{\theta \mid (\theta X^k, Y^k) \in \mathcal{T}\}$$
$$\text{output:} \quad \max\{\phi \mid (X^k, \phi Y^k) \in \mathcal{T}\}$$

See M. J. FARRELL (1957), The measurement of productive efficiency, *Journal of the Royal Statistical Society, Series A*, **120(3)**, 253–290.

Additive: This measure quantifies the maximal sum of *absolute* improvements (input reduction/output increase measured in "slacks"). It has a price interpretation (as difference between actual and maximal profit) and indicates Koopmans efficiency but it isn't invariant with respect to units of measurement.

$$\text{non-oriented:} \quad \max\left\{\sum_i s_i + \sum_j t_j \mid (X^k - s, Y^k + t) \in \mathcal{T},\, (s,t) \geqq 0\right\}$$
$$\text{input:} \quad \max\left\{\sum_i s_i \mid (X^k - s, Y^k) \in \mathcal{T},\, s \geqq 0\right\}$$
$$\text{output:} \quad \max\left\{\sum_j t_j \mid (X^k, Y^k + t) \in \mathcal{T},\, t \geqq 0\right\}$$

See A. CHARNES, W. W. COOPER, B. GOLANY, L. SEIFORD, and J. STUTZ (1985), Foundations of Data Envelopment Analysis for Pareto-Koopmans efficient empirical production functions, *Journal of Econometrics*, **30**, 91–107.

---

[1] A more detailed overview is given (in German) in H. SCHEEL (2000), *Effizienzmaße der Data Envelopment Analysis*, Gabler, Wiesbaden.

If you want to compute a *weighted* objective function $\sum_i w_i \cdot s_i$ you can do this by preprocessing the data; e. g. you may multiply each input/output i by the corresponding w_i.

maxAverage: This measure (a.k.a. Färe-Lovell or Russell or SBM measure) quantifies the maximal average of *relative* improvements (input reduction/output increase measured in percentages of the current level). It has no straightforward price interpretation but it is both an indicator for Koopmans efficiency (for positive data) and units invariant. The symbol $\odot$ denotes the componentwise product of two vectors, i. e.

$$((1-\theta)\odot X^k, (1+\phi)\odot Y^k) := ((1+\theta_1)X_1^k, \ldots, (1+\theta_m)X_m^k; (1+\phi_1)Y_1^k, \ldots, (1+\phi_n)Y_n^k).$$

$$\text{non-or.:} \quad \max\left\{\frac{\sum_{i:X_i^k>0}\theta_i + \sum_{j:Y_j^k>0}\phi_j}{\sum_{i:X_i^k>0}1 + \sum_{j:Y_j^k>0}1} \ \middle| \ \begin{matrix}(\theta,\phi) \geq 0, \\ ((1-\theta)\odot X^k, (1+\phi)\odot Y^k) \in \mathcal{T}\end{matrix}\right\}$$

$$\text{input:} \quad \min\left\{\frac{\sum_{i:X_i^k>0}\theta_i}{\sum_{i:X_i^k>0}1} \ \middle| \ (\theta\odot X^k, Y^k) \in \mathcal{T},\ \theta \leqq \vec{1}\right\}$$

$$\text{output:} \quad \max\left\{\frac{\sum_{j:Y_j^k>0}\phi_j}{\sum_{j:Y_j^k>0}1} \ \middle| \ (X^k, \phi\odot Y^k) \in \mathcal{T},\ \phi \geqq \vec{1}\right\}$$

See R. FÄRE and C. A. K. LOVELL (1978), Measuring the technical efficiency of production, *Journal of Economic Theory*, **19**, 150–162.

minAverage: This measure quantifies the *minimal* average of relative improvements which is *necessary* to become weakly efficient. (Weak efficiency means there does not exist a point in the technology set which is better in every input and output. We denote the weakly efficient subset of $\mathcal{T}$ by $\partial\mathcal{T}$.) Notice that for a weakly efficient point an arbitrary small improvement suffices to become Koopmans efficient whence the minAverage measure also quantifies the infimum average of improvements which is necessary to become Koopmans efficient.

It has neither a straightforward price interpretation nor is it an indicator for Koopmans efficiency but it is units invariant.

$$\text{non-or.:} \quad \min\left\{\frac{\sum_{i:X_i^k>0}\theta_i + \sum_{j:Y_j^k>0}\phi_j}{\sum_{i:X_i^k>0}1 + \sum_{j:Y_j^k>0}1} \ \middle| \ \begin{matrix}(\theta,\phi) \geq 0, \\ ((1-\theta)\odot X^k, (1+\phi)\odot Y^k) \in \partial\mathcal{T}\end{matrix}\right\}$$

$$\text{input:} \quad \max\left\{\frac{\sum_{i:X_i^k>0}\theta_i}{\sum_{i:X_i^k>0}1} \ \middle| \ (\theta\odot X^k, Y^k) \in \partial\mathcal{T},\ \theta \leqq \vec{1}\right\}$$

$$\text{output:} \quad \min\left\{\frac{\sum_{j:Y_j^k>0}\phi_j}{\sum_{j:Y_j^k>0}1} \ \middle| \ (X^k, \phi\odot Y^k) \in \partial\mathcal{T},\ \phi \geqq \vec{1}\right\}$$

This measure is based on ideas in A. CHARNES, J. J. ROUSSEAU, and J. H. SEMPLE (1996), Sensitivity and Stability of Efficiency Classifications in Data Envelopment Analysis, *The Journal of Productivity Analysis*, **7**, 5–18. See also W. BRIEC (1999), Hölder distance function and measurement of technical efficiency, *Journal of Productivity Analysis*, **11**, 111–132.

Superefficiency

If you choose a radial distance then EMS allows you to compute so called "superefficiency" scores by checking the box. For inefficient DMUs the superefficiency score coincides with the standard score defined above. For efficient DMUs a score is computed which indicates the maximal radial change which is feasible such that the DMU remains efficient. Formally, it is defined like the standard score but the DMU under evaluation is excluded from the constraints (i. e. the definition of the technology set). See P. ANDERSEN and N. C. PETERSEN (1993), A Procedure for Ranking Efficient Units in Data Envelopment Analysis, *Management Science*, **39**, 1261–1264.

---

big

If you have chosen the superefficiency model, then in the results table a score = `big` may appear. This means that the DMU remains efficient under *arbitrary large* increased inputs (input oriented) or decreased outputs (output oriented), respectively.

Restrict weights

If you have loaded weights restrictions data, you can check this box to incorporate the weights restrictions in the model. (If the box is not checked, then the weights restrictions will be ignored.)

5.4 Advanced modeling options

When you have opened the `Run model` dialog (`Ctrl`+M or Menu `DEA` → `Run model`) then you may specify some advanced models in `Options` which are described in the following paragraphs.

Evaluation
Technology

You may specify selections of DMUs which should be computed (`Evaluation`) and which should be used for building the envelopment (`Technology`). This allows you to compute *program efficiency*. I. e., for each DMU selected in `Evaluation` a score is computed

constrained by the DMUs selected in `Technology`. The lists allow selections of multiple entries via *Ctrl+click* and *Shift+click*.

| Window Analysis |
| Malmquist |

If you have panel data sorted by periods $T = 0, ..., t$ i.e. the first column of the data file looks like

<div align="center">

DMU 1 T0
DMU 2 T0
$\vdots$
DMU n T0
DMU 1 T1
$\vdots$
DMU n T1
$\vdots$
DMU 1 Tt
$\vdots$
DMU n Tt

</div>

then EMS supports computation of Window Analysis and Malmquist indices.

For Window Analysis you have to specify the number of periods and the window width.

For Malmquist indices you have to specify the number of periods. EMS computes then scores E(t)-T(t+1), i.e. the DMUs of period t are evaluated with respect to the technology built by the DMUs in period t+1. The scores E(t)-T(t) can be computed by running a Window Analysis with window width = 1. Dividing these scores is left to your spreadsheet. See the `Malmquist` sheet in `example.xls` for details.

6   Results

If computations are finished, EMS will display the results in a table. The window caption tells which model was computed, e.g. `example.xls_CRS_RAD_IN_WR-example` contains the results of a DEA model based on the input output data file `example.xls` with constant returns to scale, radial distance, input orientation, weights restrictions with restriction matrix stored in `example.xls`.

The result table contains: (recall that the number of decimals to display can be modified in Menu `DEA` → `Format`)

DMU name. An additional {X} indicates that this DMU was excluded from building the technology as specified in `Technology`. A DMU name without score indicates that this DMU built the technology but was not evaluated as specified in `Evaluation`.

The efficiency score as defined above.

the weights (shadow prices) {W} or virtual inputs/outputs {V} as selected in Menu `DEA` → `Format`,

benchmarks:

- for inefficient DMU: the reference DMUs with corresponding intensities (the "lambdas") in brackets
- for efficient DMU: the number of inefficient DMUs which have chosen the DMU as Benchmark,

slacks {S} or factors {F}. Depending of the chosen distance, for radial and additive measures the slacks are displayed. For the minAverage and maxAverage measures the factors (i.e. the θ_i, ϕ_j as defined above) are displayed. In addition, for the minAverage measure slacks are displayed for those inputs and outputs with factors = 1 (or 0 for non-oriented measure).

For Nonconvex (FDH) models, instead of the weights for each DMU the number of dominated and dominating DMUs and lists of these DMUs are displayed.

```
Copy
Save
```

The result table can be copied to your spreadsheet via the Windows clipboard: Use Menu `Edit` → `Copy` (*Ctrl*+C) for copying your current selection or `Copy All` (*Ctrl*+A) for the whole results table.

The result table can be saved as an ASCII file (Menu `File` → `Save` or `Save As`).

7   Acknowledgements

Thanks to BPMPD's author Csaba Mészáros for the kind support during the development of EMS. Thanks to Laurens Cherchye for valuable discussions.

8   Disclaimer

The author of the program described here accepts no responsibility for damages resulting from the use of this software and makes no warranty, either express or implied, including, but not limited to, any implied warranty of fitness for a particular purpose. The software is provided as it is, and you, its user, assume all risks when using it.