DEA-Solver-Pro Newsletter

No. 15
(July 5, 2015)

FAQ

Q: How can we deal with big data set comprised of several thousand DMUs?

A: You can divide them into several groups consisting of a few hundred DMUs and solve each group for finding efficient DMUs in the group. Then, merge the efficient DMUs into one set. Solve the set and find the final efficient DMUs. Using the finalist, re-evaluate each group.

Version 13.0 SBM_Max

In Version 13.0, we replaced *SBM_Variation* models in the earlier version by *SBM_Max* models. The SBM models usually report the worst efficiency scores for inefficient DMUs. This means that the projected point is the farthest one on the associated efficient frontier. In contrast, *SBM_Max* models look for the nearest point on the associated efficient frontier. Hence, the efficiency score is, in a sense, maximized as contrasted to the ordinary *SBM* (*Min*) models. This indicates that we can attain an efficient status with less input reductions and less output expansions than the ordinary SBM (Min) models. The relationship among SBM-Min, SBM-Max and CCR models can be depicted by the figure below. The new SBM_Max models attain more accurate nearest point on the efficient frontiers than the SBM_Variation in the previous version, whereas the computational time increases depending on the number of efficient DMUs. We can say that the projection by SBM_Max model represents a practical “KAIZEN” (Improvement) by DEA. This model includes input-, output- and non-oriented versions under the constant and variable returns-to-scale assumptions.

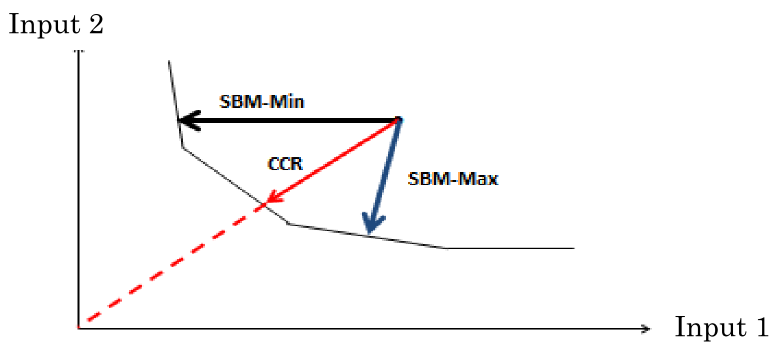


Figure 13.1

Version 12.0: Directional Distance Model in DEA

In Version 12.0, we introduced *Directional Distance (Function) Model* that computes the efficiency of DMUs along the given direction. We offer two types of efficiency, i.e., ordinary and super efficiency.
The Directional Distance model (DD model) has three variations: (1) Input- (2) Output- and (3) Non-oriented. Each variation has two returns-to-scale (RTS): (a) CRS (Constant RTS) and (b) VRS (Variable RTS).
The DD model deals with two kinds of input and three kinds of output as follows.

(a) Directional (Ordinary) inputs (I)

The heading of this class is (I) and its value is denoted by $x_{ij}^{(I)}\ (i=1,\ldots,I;\ j=1,\ldots,n)$ where $i=1,\ldots,I$ corresponds to input and $j=1,\ldots,n$ indicates DMU. The component of directional vector is defined as $d_{ij}^{(I)}=x_{ij}^{(I)}$.

(b) Non-directional inputs (IN)

The heading of this class is (IN) and denoted by $x_{ij}^{(IN)}\ (i=1,\ldots,IN;\ j=1,\ldots,n)$ where IN is the number of (IN) inputs. The component of directional vector is defined as $d_{ij}^{(IN)}=0$. This class has no directional effect on the efficiency score, but associates with it through the intensity vector. This class is the secondary input and can be empty.

(c) Directional (Good) output (O)

The heading of this class is (O) and denoted by $y_{ij}^{(O)}\ (i=1,\ldots,\mathrm{O};\ j=1,\ldots,n)$ where O is the number of good outputs. The component of directional vector is defined as $d_{ij}^{(O)}=y_{ij}^{(O)}$.

(d) Non-directional outputs (ON)

The heading of this class is (ON) and denoted by $y_{ij}^{(ON)}\ (i=1,\ldots,\mathrm{ON};\ j=1,\ldots,n)$ where ON is the number of non-directional outputs. The component of directional vector is defined as $d_{ij}^{(ON)}=0$. This class has no directional effect on the efficiency score, but associates with it through the intensity vector. This class is the secondary output and can be empty.

(e) Undesirable outputs (OBad)

The heading of this class is (OBad) and denoted by $y_{ij}^{(OBad)}\ (i=1,\ldots,\mathrm{OBad};\ j=1,\ldots,n)$ where OBad is the number of undesirable outputs. The component of directional vector is defined as $d_{ij}^{(OBad)}=y_{ij}^{(OBad)}$. This class can be empty.

Version 11: Resampling

In Version 11.0, we introduced a new model called "Resampling in DEA". This model deals with data variations and gives confidence intervals of DEA scores. Input/output values are subject to change for several reasons, e.g., measurement errors, hysteretic factors, arbitrariness and so on. Furthermore, these variations differ in their input/output items and their decision-making units (DMU). Hence, DEA efficiency scores need to be examined by considering these factors. Resampling based on these variations is necessary for gauging the confidence interval of DEA scores. We propose four resampling models. The first one assumes downside and upside measurement error rates for each input/output, which are common to all DMUs. We resample data following the triangular distribution that the downside and upside errors indicate around the observed data. The second model evaluates the downside and upside error rates from historical data. The third model utilizes historical data, e.g., past-present, for estimating data variations, imposing chronological order weights which are supplied by Lucas series (a variant of Fibonacci series). This model provides different downside and upside error rates for each DMU. The last one deals with future prospects. This model aims at forecasting the future efficiency score and its confidence interval for each DMU. Table 11.1 exhibits a sample of forecast DEA scores and their confidence intervals, while Figure 11.1 shows 95% confidence interval of forecast scores.

Table 11.1: Forecast DEA score and confidence interval

DMU	Forecast	97.50%	90%	80%	75%	50%	25%	20%	10%	2.50%
H1	0.946	1.122	1.092	1.062	1.036	0.925	0.880	0.870	0.849	0.818
H2	1.043	1.128	1.098	1.079	1.073	1.049	0.906	0.883	0.845	0.815
H3	0.708	0.743	0.722	0.709	0.704	0.683	0.662	0.657	0.645	0.625
H4	0.809	0.828	0.809	0.796	0.792	0.769	0.747	0.742	0.728	0.708
H5	0.719	0.748	0.727	0.714	0.709	0.689	0.668	0.663	0.651	0.633
H6	1.067	1.185	1.146	1.124	1.115	1.084	1.055	1.049	1.033	0.993
H7	0.900	0.918	0.894	0.879	0.874	0.849	0.826	0.820	0.805	0.782
H8	0.795	0.877	0.835	0.809	0.798	0.762	0.730	0.722	0.705	0.680
H9	0.689	0.736	0.707	0.689	0.681	0.655	0.630	0.623	0.606	0.582
H10	0.801	1.065	0.981	0.914	0.891	0.791	0.748	0.740	0.720	0.689
H11	0.866	1.010	0.947	0.914	0.902	0.861	0.826	0.819	0.798	0.771
H12	1.013	1.014	0.992	0.961	0.951	0.911	0.878	0.870	0.850	0.822

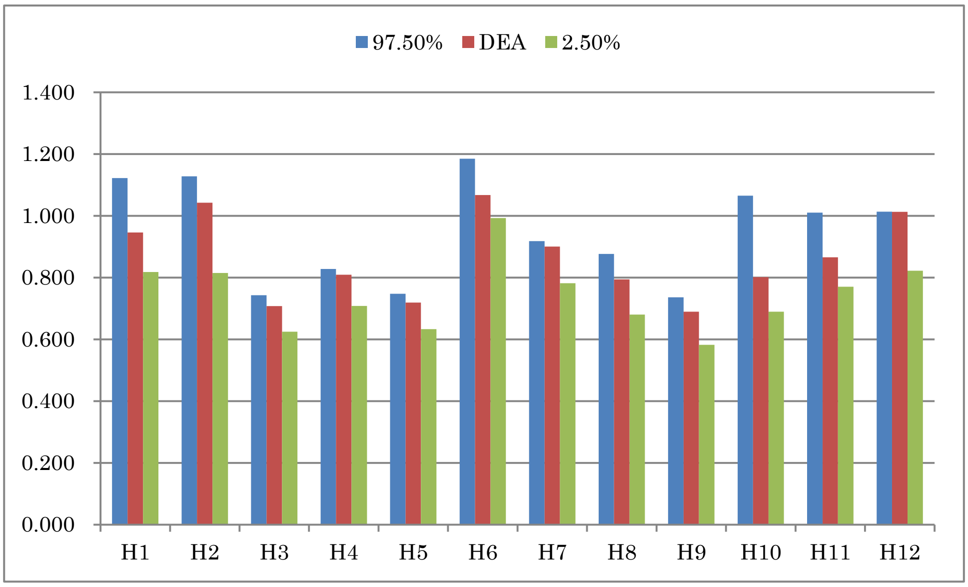


Figure 11.1: Forecast DEA score and 95% confidence interval

Version 10: Non-convex DEA

In Version 10.0, we introduced the “Non-convex DEA” model. In DEA, we are often puzzled by the large difference between the constant-returns-scale (CRS) and variable returns-to-scale (VRS) scores, and by the convexity production set syndrome in spite of the S-shaped curve often observed in many real data sets. See Figure 10.1. In this model, we propose a solution to these problems.

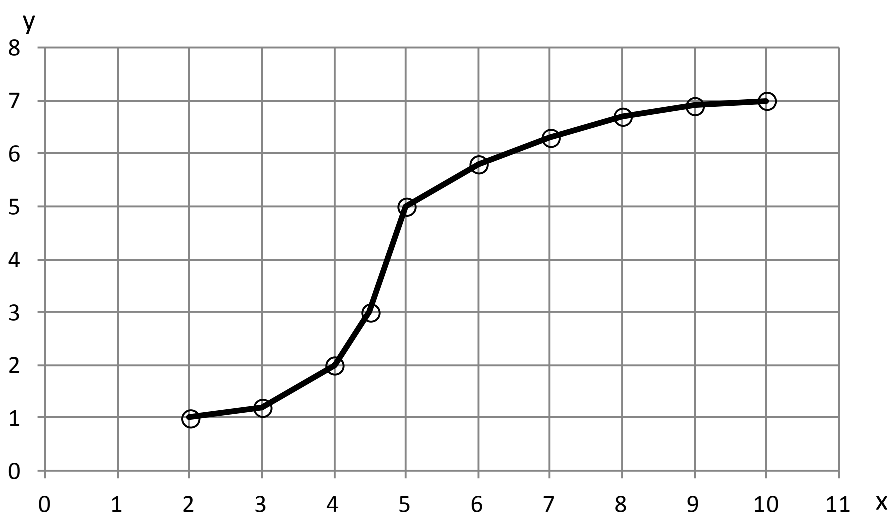


Figure 10.1: S-shaped data

Version 9: Dynamic and Network DEA

In Version 9.0, we introduced a model called “DNSBM” (dynamic SBM with network structure). We have already introduced the network SBM (NSBM) and the dynamic SBM (DSBM) models separately. Hence, this model is a composite of these two models. Vertically, we deal with multiple divisions connected by links of network structure within each period and, horizontally, we combine the network structure by means of carry-over activities between two succeeding periods. See Figure 9.1 below. This model can evaluate (1) the overall efficiency over the observed entire period, (2) dynamic change of period (term) efficiency (3) dynamic change of divisional efficiency and (4) divisional Malmquist index. The model can be implemented in input-, output- or non-(both) oriented forms under the CRS or VRS assumptions on the production possibility set. We can impose the initial condition on the carry-overs.

We can formally deal with bad carry-overs, such as nonperforming loans and dead stock, in this model. We modified the data formats of Dynamic SBM (DSBM) and Network SBM (NSBM) so that they are compatible with DNSBM model. Please see User’s Guide Version 10.0.

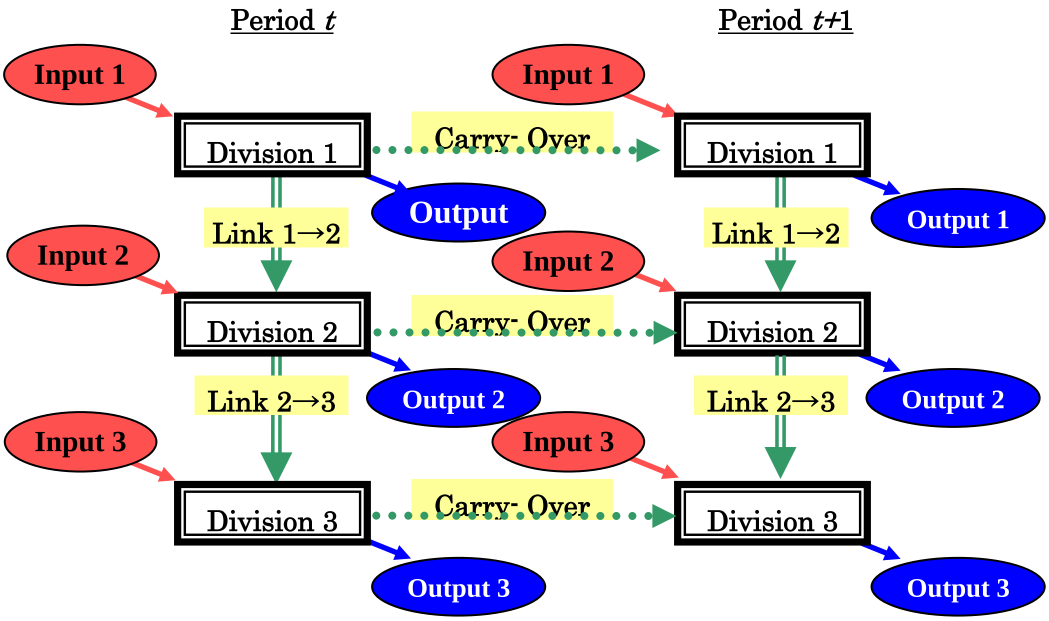


Figure 9.1: Dynamic model with network structure

Version 8: EBM

In Version 8.0 we introduced a model called “EBM” (epsilon-based measure of efficiency) for measuring technical efficiency. DEA is a data driven tool for measuring efficiency of decision making units (DMU) and shows a sharp contrast to so-called “parametric methods” such as stochastic frontier analysis (SFA). The latter methods assume specific production function forms to be identified. This assumption is not so reasonable in several instances and aspects. Since DEA can deal with multiple inputs vs. multiple outputs relations in a single framework, it is becoming a method of choice for efficiency evaluation. However, DEA has several shortcomings to be explored further. In DEA, we have two measures of technical efficiency with different characteristics: radial and non-radial. Historically, the radial measure, represented by the CCR model was the first DEA model, whereas the non-radial model, represented by the SBM model was a latecomer. For instance, in the input-oriented case, the CCR deals mainly with proportionate reduction of input resources. In other words, if the organisational unit under study has two inputs, this model aims at obtaining the maximum rate of reduction with the same proportion, i.e. a radial contraction in the two inputs that can produce the current outputs. In contrast, the non-radial models put aside the assumption of proportionate contraction in inputs and aim at obtaining maximum rates of reduction in inputs that may discard varying proportions of original input resources.

In this EBM, we propose a composite model which combines both models in a unified framework. This model has two parameters: one scalar and one vector. In order to determine these two parameters, we introduce a new affinity index associated with inputs or outputs. We apply the principal component analysis (PCA) to thus defined affinity matrix. These two parameters are utilized to unify radial and non-radial models in a single model.

Version 7: Dynamic DEA

In Version 7.0, we reinforced our collection of DEA models by the addition of Dynamic DEA models. In DEA, there are several methods for measuring efficiency changes over time, e.g. the Window analysis and the Malmquist index. However, they usually neglect carry-over activities between two consecutive terms and only focus on the separate period independently aiming local optimization in a single period, even if these models can take into account the time change effect. In the actual business world, a long time planning and investment is a subject of great concern. For these cases, single period optimization model is not suitable for performance evaluation. To cope with long time point of view, the dynamic DEA model incorporates carry-over activities into the model as depicted below and enables us to measure period specific efficiency based on the long time optimization during the whole period.

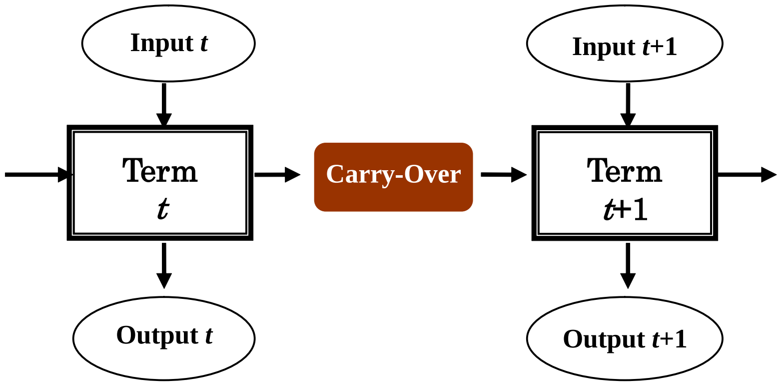
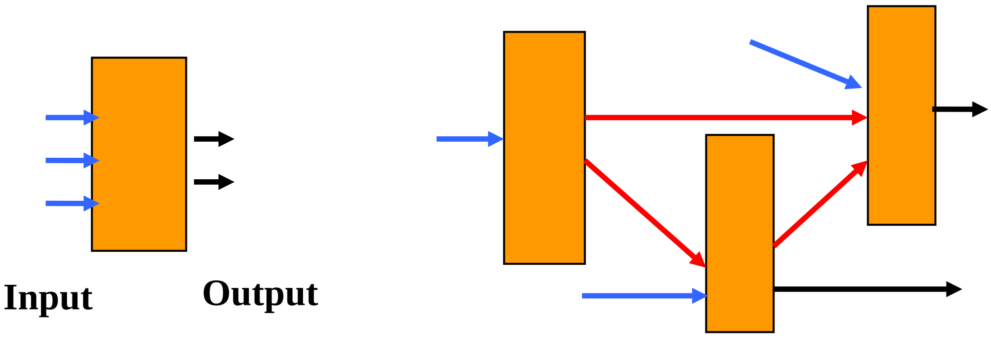


Figure 7.1: Dynamic model

We enforced Network SBM models by adding the uniqueness check of the divisional efficiencies.

Version 6: Network DEA

In Version 6.0 we added Network DEA (Network SBM) models. Network DEA intends to fortify DEA models in the following way. Traditional DEA models evaluate the relative efficiency of DMUs regarding multiple-inputs vs. multiple-outputs correspondences. Figure below (left) exhibits such traditional DEA models. However, looking into the inside of the DMU, we often find inter-related divisions as exemplified in Figure below (right). In this example, an output from Division 1 (Div1) comes to Division 2 (Div2) as an input. The three red lines are intermediate products. We call then as "link." Traditional DEA models cannot deal with links formally, since every activity should belong to inputs or outputs, but not both. These inner structures form a network, and the Network DEA treats these inter-related organizations formally. Using this model, we can identify the relative efficiency of each division as well as the overall efficiency of the DMU. Network DEA serves to clarify the internally-related efficiencies which are usually regarded as "black box" and have not been investigated so far.



Version 5: Hybrid and Weighted SBM

In Version 5.0, we added our collection of DEA models by the Hybrid model and the Weighted SBM model, and modified the Non-separable Outputs model in the Undesirable Outputs cluster. We also added worksheets “Decomposition” to the SBM-family and Non-separable Output models. This includes the decomposition of efficiency score into inefficiency of each input/output factor.

The Hybrid model deals with radial and non-radial inputs/outputs in a unified framework. The Weighted SBM model puts weights to slacks in accordance with their importance to management.

Version 4: Undesirable Outputs

In Version 4.1, we enriched our collection of DEA models by the addition of Undesirable Outputs models. In accordance with the global environmental conservation awareness, undesirable outputs of production and social activities, e.g., air pollutants and hazardous waste, have been widely recognized as societal evils. Thus, development of technologies with less undesirable outputs is an important subject of concern in every area of production. DEA usually assumes that producing more outputs relative to less input resources is a criterion of efficiency. In the presence of undesirable outputs, however, technologies with more good (desirable) outputs and less bad (undesirable) outputs relative to less input resources should be recognized as efficient. The Undesirable Model deals with this problem. Please see Newsletter 5.

All DEA models can be classified into four types: (1) Radial, (2) Non-Radial and Oriented, (3) Non-Radial and Non-Oriented and (4) Composite of Radial and Non-radial. 'Radial' means that a proportionate change of input/output values is the main concern and hence it neglects the existence of slacks (input excesses and output shortfalls remaining in the model) as secondary or freely disposable, whereas 'Non-Radial' deals with slacks directly and does not stick to a proportionate change of input/output. 'Oriented' indicates the input or output orientation in evaluating efficiency, i.e., the main target of evaluation is either input reduction or output expansion. For example, input oriented models first aim to reduce input resources to the efficient frontier as far as possible, and then to enlarge output products as the second objective. 'Non-Oriented' models deal with input reduction and output expansion at the same time.

They are classified into the four categories as displayed below.

Category	Cluster or Model
Radial	CCR, BCC, IRS, DRS, AR, ARG, NCN, NDSC, BND, CAT, SYS, Bilateral, Scale Elasticity, Congestion, Window, Malmquist-Radial, FDH, NonConvex-Radial, Resampling –SuperRadial DD-I(O)-C(V), SuperDD-I(O)-C(V), DD-C(V), SuperDD-C(V)
Non-Radial and Oriented	SBM-Oriented, Super-efficiency-Oriented, Malmquist, NetworkSBM(Oriented), DynamicSBM(Oriented), DynamicNetworkSBM(Oriented), NonConvex-SBM-I(O), Resampling-SuperSBM(Oriented), SBM Max(Oriented)
Non-Radial and Non-Oriented	Cost, New-Cost, Revenue, New-Revenue, Profit, New-Profit, Ratio, SBM-NonOriented, Super-SBM-NonOriented, Malmquist-C (V, GRS), Undesirable Outputs, Weighted SBM , NetworkSBM(NonOriented), DynamicSBM(NonOriented), Bilateral, DynamicNetworkSBM(NonOriented), NonConvex-SBM-NonOriented, Resampling-SuperSBM(NonOriented), SBM Max (NonOriented)
Radial and Non-Radial	Hybrid, EBM

The following 51 clusters are included.

No.	Cluster	Model
1	CCR	CCR-I, CCR－O
2	BCC	BCC-I, BCC-O
3	IRS	IRS-I, IRS-O
4	DRS	DRS-I, DRS-O
5	GRS	GRS-I, GRS-O
6	AR (assurance region)	AR-I-C, AR-I-V, AR-I-GRS, AR-O-C, AR-O-V, AR-O-GRS
7	ARG (assurance region global)	ARG-I-C, ARG-I-V, ARG-I-GRS, ARG-O-C, ARG-O-V, ARG-O-GRS
8	NCN (non-controllable)	NCN-I-C, NCN-I-V, NCN-O-C, NCN-O-V
9	NDSC (non-discretionary)	NDSC-I-C, NDSC-I-V, NDSC-I-GRS, NDSC-O-C, NDSC-O-V, NDSC-O-GRS
10	BND (bounded variable)	BND-I-C, BND-I-V, BND-I-GRS, BND-O-C, BND-O-V, BND-O-GRS
11	CAT (categorical variable)	CAT-I-C, CAT-I-V, CAT-O-C, CAT-O-V
12	SYS (different systems)	SYS-I-C, SYS-I-V, SYS-O-C, SYS-O-V
13	SBM-Oriented (Slacks-based Measure)	SBM-I-C, SBM-I-V, SBM-I-GRS, SBM-O-C, SBM-O-V, SBM-O-GRS, SBM-AR-I-C, SBM-AR-I-V, SBM-AR-O-C, SBM-AR-O-V
14	SBM-Non Oriented	SBM-C, SBM-V, SBM-GRS, SBM-AR-C, SBM-AR-V
15	Weighted SBM	WeightedSBM-C, WeightedSBM-V, WeightedSBM-I-C, WeightedSBM-I-V, WeightedSBM-O-C, WeightedSBM-O-V
16	Super-SBM-Oriented	Super-SBM-I-C, Super-SBM-I-V, Super-SBM-I-GRS, Super-SBM-O-C, Super-SBM-O-V, Super-SBM-O-GRS
17	Super-SBM-NonOriented	Super-SBM-C, Super-SBM-V, Super-SBM-GRS
18	Super-Radial	Super-CCR-I, Super-CCR-O, Super-BCC-I, Super-BCC-O
19	Cost	Cost-C, Cost-V, Cost-GRS
20	New-Cost	New-Cost-C, New-Cost-V, New-Cost-GRS
21	Revenue	Revenue-C, Revenue-V, Revenue-GRS
22	New-Revenue	New-Revenue-C, New-Revenue-V, New-Revenue-GRS
23	Profit	Profit-C, Profit-V, Profit-GRS
24	New-Profit	New-Profit-C, New-Profit-V, New-Profit-GRS
25	Ratio (Revenue/Cost)	Ratio-C, Ratio-V
26	Bilateral	Bilateral-CCR-I, Bilateral-BCC-I, Bilateral-SBM-C, Bilateral-SBM-V
27	Window	Window-I-C, Window-I-V, Window-I-GRS, Window-O-C, Window-O-V, Window-O-GRS
28	FDH	FDH
29	Malmquist-NonRadial	Malmquist-I-C, Malmquist-I-V, Malmquist-I-GRS, Malmquist-O-C, Malmquist-O-V, Malmquist-O-GRS, Malmquist-C, Malmquist-V, Malmquist-GRS
30	Malmquist-Radial	Malmquist-Radial-I-C, Malmquist-Radial-I-V, Malmquist-Radial-I-GRS, Malmquist-Radial-O-C, Malmquist-Radial-O-V, Malmquist-Radial-O-GRS
31	Scale Elasticity	Elasticity-I, Elasticity-O
32	Congestion	Congestion
33	Undesirable outputs	BadOutput-C, BadOutput-V, BadOutput-GRS,

		NonSeparable-C, NonSeparable-V, NonSeparable-GRS
34	**Hybrid**	**Hybrid-C, Hybrid-V, Hybrid-I-C, Hybrid-I-V, Hybrid-O-C, Hybrid-O-V**
35	**Network DEA Oriented**	**NetworkSBM-I(O)-C(V)**
36	**Network DEA Non-Oriented**	**NetwrokSBM-C(V)**
37	**Dynamic DEA Oriented**	**DynamicSBM-I(O)-C(V)**
38	**Dynamic DEA Non-oriented**	**DynamicSBM-C(V)**
39	**EBM Oriented**	**EBM-I(O)-C(V)**
40	**EBM Non-Oriented**	**EBM-C(V)**
41	**DynamicNetworkSBM Oriented**	**DNSBM-I-C(V), DNSBM-O-C(V)**
42	**DynamicNetworkSBM NonOriented**	**DNSBM-C(V)**
43	**NonConvex-SBM**	**NonConvex-SBM-I(O),-NonOriented**
44	**NonConvex-Radial**	**NonConvex-Radial-I(O)**
45	**Rsampling-Triangular**	**Resampling-(Super)SBM, Resampling-(Super)Radial**
46	**Resampling-Triangular-Historical**	**Resampling-(Super)SBM, Resampling-(Super)Radial**
47	**Resampling-PastPresent**	**Resampling-(Super)SBM, Resampling-(Super)Radial**
48	**Resampling-PastPresentFuture**	**Resampling-(Super)SBM, Resampling-(Super)Radial**
49	**Directional Distance Oriented**	**DD-I(O)-C(V), SuperDD-I(O)-C(V)**
50	**Directional Distance Non-oriented**	**DD-C(V), SuperDD-C(V)**
51	**SBM_Max**	**SBM_Max-I(O)-C(V), -C(V)**

The meanings of the extensions -C, -V and –GRS are as follows. Every DEA model assumes a returns to scale (RTS) characteristics that is represneted by the range (*L, U*) of the sum of the intensity vector λ . The constant RTS (-C) corresponds to (*L*=0, *U*=infinity) and the variable RTS (-V) to (*L*=1, *U*=1), respectively.

In the models with the extension GRS, we have to supply L and U from keyboard, the defaults being *L*=0.8 and *U*=1.2. The increasing RTS (IRS) corresponds to (*L*=1, *U*=infinity) and the decreasing RTS (DRS) to (*L*=0, *U*=1), respectively. It is recommended to try several sets of (*L, U*) in order to identify how the RTS chracteristics exerts an influence on the efficiency score. In the Non-Oriented EBM models (EBM-C and EBM-V), we must supply *L* and *U* from the keyboard, the defaults being *L*=*U*=1.

In the Resampling models, the number of replicas must be supplied through the keyboard.