

HOW TO SELECT AN EPSILON TO APPLY KOUROSH AND ARASH MODEL IN DATA ENVELOPMENT ANALYSIS

DARIUSH KHEZRIMOTLAGH

Department of Applied Statistics, Faculty of Economics and Administration, University of Malaya, Kuala Lumpur, Malaysia

ABSTRACT

This paper demonstrates how an appropriate epsilon value can be selected when Kourosh and Arash Model (KAM) is applied in order to estimate production frontiers as well as simultaneously rank and benchmark Decision Making Units (DMUs). KAM was recently proposed to improve the capabilities of Data Envelopment Analysis (DEA) in order to measure the performance evaluation of homogenous DMUs inclusive real and integer values. The paper also illustrates that the selection of epsilon is logically allowed even if the DMUs' data are exact. A non-linear programming model is also proposed to find the optimum value of score while a DMU is under evaluation.

Keywords: DEA; KAM; Efficiency; Ranking; Benchmarking.

INTRODUCTION

Data Envelopment Analysis (DEA) was proposed by Charnes *et al.* (1978), and has been applied in many area of economics, management, industrial engineering, business, marketing and so on to measure the performance evaluation of homogenous Decision Making Units (DMUs). Recently Khezrimotlagh *et al.* (2013a) proposed a new model in DEA, called Kourosh and Arash Model (KAM), to improve the advantages of DEA to simultaneously benchmark both technically efficient and inefficient DMUs. KAM unlike current DEA models provides a methodology based on an introduced epsilon which is able to measure the efficiency score of DMUs where the weights are available or unknown. In this paper, it is illustrated how an epsilon can be selected while KAM is applied.

KOUROSH AND ARASH MODEL (KAM)

Suppose that there are n DMUs ($DMU_i, i = 1, 2, \dots, n$) with m non-negative inputs ($x_{ij}, j = 1, 2, \dots, m$) and p non-negative outputs ($y_{ik}, k = 1, 2, \dots, p$), such that, at least one of the inputs and one of the outputs of each DMU are not zero. Assume that for every i there is a j such that $x_{ij} \neq 0$ and also for every i there is a k such that $y_{ik} \neq 0$. A Production Possibility Set (PPS) with Variable Returns to Scale (VRS) technology proposed by Banker *et al.* (1984) is shown with T_V . The geometric locus of points in \mathbb{R}_+^{m+p} which their distances of the T_V frontier are not greater than $\varepsilon (\geq 0)$ is called an efficient hyper-tape, and denoted with $TT_C^{+\varepsilon}$. An ε -higher PPS ($\varepsilon \in \mathbb{R}_+^{m+p}$) in KAM-VRS, denoted by $T_V^{+\varepsilon}$ is the summation of T_V and $TT_V^{+\varepsilon}$. The ε -KAM on $T_V^{+\varepsilon}$ is as follows where DMU_l for $l = 1, 2, \dots, n$ is under evaluation:

$$\max \sum_{j=1}^m w_{lj}^- s_j^- + \sum_{k=1}^p w_{lk}^+ s_{lk}^+,$$

Subject to

$$\sum_{i=1}^n \lambda_i x_{ij} + s_{lj}^- = x_{lj} + \varepsilon_{lj}^-, \text{ for } j = 1, 2, \dots, m,$$

$$\sum_{i=1}^n \lambda_i y_{ik} - s_{lk}^+ = y_{lk} - \varepsilon_{lk}^+, \text{ for } k = 1, 2, \dots, p,$$

$$\sum_{i=1}^n \lambda_{li} = 1, \quad x_{lj} - s_{lj}^- \geq 0, \quad \text{for } j = 1, 2, \dots, m, \quad y_{lk} + s_{lk}^+ - 2\varepsilon_{lk}^+ \geq 0, \quad \text{for } k = 1, 2, \dots, p,$$

$$\lambda_{li} \geq 0, \quad \text{for } i = 1, 2, \dots, n, \quad s_{lj}^- \geq 0, \quad \text{for } j = 1, 2, \dots, m, \quad s_{lk}^+ \geq 0, \quad \text{for } k = 1, 2, \dots, p.$$

The outcomes of KAM are given by:

The best technical efficient target with ϵ degree of freedom (ϵ -DF) and the best technical efficiency score with ϵ -DF can respectively be expressed as:

$$\begin{cases} x_{lj}^* = x_{lj} - s_{lj}^- + \varepsilon_{lj}^-, & \text{for } j = 1, 2, \dots, m, \\ y_{lk}^* = y_{lk} + s_{lk}^+ - \varepsilon_{lk}^+, & \text{for } k = 1, 2, \dots, p. \end{cases}$$

$$KA_{\varepsilon}^{*l} = \frac{\sum_{k=1}^p w_{lk}^+ y_{lk} / \sum_{j=1}^m w_{lj}^- x_{lj}}{\sum_{k=1}^p w_{lk}^+ y_{lk}^* / \sum_{j=1}^m w_{lj}^- x_{lj}^*}.$$

The highest efficient target with ϵ -DF and the lowest efficiency score with ϵ -DF are respectively represented by:

$$\begin{cases} \widehat{x}_{lj}^* = x_{lj} - s_{lj}^-*, & \text{for } j = 1, 2, \dots, m, \\ \widehat{y}_{lk}^* = y_{lk} + s_{lk}^+*, & \text{for } k = 1, 2, \dots, p. \end{cases}$$

$$\widehat{KA}_{\varepsilon}^{*l} = \frac{\sum_{k=1}^p w_{lk}^+ y_{lk} / \sum_{j=1}^m w_{lj}^- x_{lj}}{\sum_{k=1}^p w_{lk}^+ \widehat{y}_{lk}^* / \sum_{j=1}^m w_{lj}^- \widehat{x}_{lj}^*}.$$

The lowest efficient target with ϵ -DF and the highest efficiency score with ϵ -DF are respectively depicted as:

$$\begin{cases} \widetilde{x}_{lj}^* = x_{lj} - s_{lj}^- + 2\varepsilon_{lj}^-, & \text{for } j = 1, 2, \dots, m, \\ \widetilde{y}_{lk}^* = y_{lk} + s_{lk}^+ - 2\varepsilon_{lk}^+, & \text{for } k = 1, 2, \dots, p. \end{cases}$$

$$\widetilde{KA}_{\varepsilon}^{*l} = \frac{\sum_{k=1}^p w_{lk}^+ y_{lk} / \sum_{j=1}^m w_{lj}^- x_{lj}}{\sum_{k=1}^p w_{lk}^+ \widetilde{y}_{lk}^* / \sum_{j=1}^m w_{lj}^- \widetilde{x}_{lj}^*}.$$

The best efficient region with ϵ -DF is a hyper-cube with center (x_l^*, y_l^*) and diagonal $2\left(\sum_{j=1}^m (\varepsilon_{lj}^-)^2 + \sum_{k=1}^p (\varepsilon_{lk}^+)^2\right)^{1/2}$.

APPLYING KAM

In Applying KAM, if the value of epsilon is 0, KAM is the same as the weighted Additive DEA model (ADD) suggested by Charnes et al. (1985). When epsilon has a positive value, KAM tests whether the close inefficient points to evaluated DMU has a close efficiency score as the DMU or the differences are significant, even if data are exact. KAM does not change data, but checks the scores of DMUs' neighbours.

The epsilon is a vector, i.e., $\epsilon = (\varepsilon^-, \varepsilon^+) \in \mathbb{R}_+^{m+p}$, where ε^- is $(\varepsilon_1^-, \varepsilon_2^-, \dots, \varepsilon_m^-)$ and ε^+ is $(\varepsilon_1^+, \varepsilon_2^+, \dots, \varepsilon_p^+)$. It is strongly suggested to consider the commensurate values for each component of epsilon corresponding to each variable. One of the ways to select epsilons is to define $\varepsilon_j^- = \varepsilon x_j$, and $\varepsilon_k^+ = \varepsilon y_k$,

for a suitable $\varepsilon \in \mathbb{R}_+$, $j = 1, 2, \dots, m$, and $k = 1, 2, \dots, p$. The effects on the Farrell frontier in this case are different for each evaluated DMU. In other words, when a DMU has large input/output values, the diagonal of efficient tape is greater than when it has small input/output values. The ε_{lj}^- and ε_{lk}^+ can be defined as $\varepsilon \times \min\{x_{ij}: x_{ij} \neq 0, i = 1, 2, \dots, n\}$ and $\varepsilon \times \min\{y_{ik}: y_{ik} \neq 0, i = 1, 2, \dots, n\}$, where $\varepsilon \in \mathbb{R}_+$, to have the same commensurate effects in the Farrell frontier for evaluating each DMU. In these two kind of selecting epsilons instead ε -DF it is written ε -DF.

If user only wants to select the targets which lie on the Farrell frontier, it is suggested to use the best technical efficiency score with ε -DF by selecting the best technical efficient target with ε -DF. In this case, user can remove two classes of constraints, which are $x_{lj} - s_{lj}^- \geq 0$, for $j = 1, 2, \dots, m$, and $y_{lk} + s_{lk}^+ - 2\varepsilon_{lk}^+ \geq 0$, for $k = 1, 2, \dots, p$, and can select the value of epsilon in the interval $[0, \infty)$ before applying KAM. However, when the value of diagonal of efficient tape is considered as a negligible value the benchmarking and ranking would be more significant and reasonable.

For the highest efficient target with ε -DF and the lowest efficiency score with ε -DF, user can remove the constraints $y_{lk} + s_{lk}^+ - 2\varepsilon_{lk}^+ \geq 0$, for $k = 1, 2, \dots, p$, and for the lowest efficient target with ε -DF and the highest efficiency score with ε -DF, user can remove $x_{lj} - s_{lj}^- \geq 0$, for $j = 1, 2, \dots, m$, before applying KAM. Note that, for these scores the value of epsilon should be suitable to have a feasible targets. One simple way to remove this concern is to consider the ε_{lj}^- and ε_{lk}^+ as $\varepsilon \times \min\{x_{ij}: x_{ij} \neq 0, i = 1, 2, \dots, n\}$ and $\varepsilon \times \min\{y_{ik}: y_{ik} \neq 0, i = 1, 2, \dots, n\}$, where ε is positive and small enough. Moreover, if user is interest to have all the KAM scores, none of the constraints should be removed.

Usually data has two or three decimal digits and regards to the number of variables, epsilon can be considered as one millionth or one hundred thousandth in order to have a negligible errors in. Khezrimotlagh et al. (2012b), Mohsenpour et al. (2013) and Khezrimotlagh (2014) proved that the results of Arash Model (AM), Kourosh Model (KM) and KAM are the same as the results of Cost efficiency, Revenue efficiency and Profit efficiency, respectively, when the weights of variables are known and the value of epsilon is large enough. Note that, in these statements, there are some conditions to apply KAM.

If the weights are unknown, it is suggested to define, $w_j^- = 1/x_j$, $w_k^+ = 1/y_k$, where $x_j \in \mathbb{R}_+$, $y_k \in \mathbb{R}_+$, for $j = 1, 2, \dots, m$, and $k = 1, 2, \dots, p$. If $x_j = 0$ ($y_k = 0$), for some $j = 1, 2, \dots, m$ ($k = 1, 2, \dots, p$), the weight w_j^- (w_k^+) can be defined as 1 or any other positive numbers related to the goal of DMUs as suggested by Tone (2001).

In order to review some examples, user can check the applicable examples in Khezrimotlagh et al. (2012a-f, 2013a-e). For instance, in applicable example of 19 electric utilities operating in 1975 with three inputs (labour, fuel and capital) and one output (electric power), (proposed by Färe et al. (1989), used by Simar and Wilson (1998) using bootstrapping algorithm), Khezrimotlagh et al. (2013a) used the minimum amounts of non-zero inputs and output values which were 94.00, 540.61, 182.30 and 457.20, respectively. Then they assumed that $\varepsilon = 0.00001 = 10^{-5}$ and calculated the components of ε vector as $\varepsilon_1^- = 0.0009400$, $\varepsilon_2^- = 0.0054061$, $\varepsilon_3^- = 0.0018230$ and $\varepsilon^+ = 0.0045720$, respectively, whit the diameter of efficient tape as 0.0147426. They later showed that if epsilon is considered as $\varepsilon = 0.1 =$

10^{-1} , the results are almost the same as the result when $\varepsilon = 0.00001 = 10^{-5}$, whereas the diameter of efficient tape is 147.4260. For another example, Khezrimotlagh et al. (2013b) considered 39 Spanish airports used in the study of Lozano and Gutierrez (2011) with 4 inputs and 3 outputs. The minimum amounts of non-zero inputs values were 1, 1, 1 and 1, respectively. They assumed that $\varepsilon = 0.00001 = 10^{-5}$, then the ε vector was $(10^{-5}, 10^{-5}, 10^{-5}, 10^{-5}, 0, 0, 0)$. The diagonal of efficient tape was $0.00004 = 4 \times 10^{-5}$, which was exactly negligible where inputs are restricted in the set of integer numbers. Even if ε is defined as $0.1 = 10^{-1}$, the diameter of efficient tape is 0.4 only, which is still negligible where inputs are restricted in the set of integer numbers. With these scenarios they depicted the significant results of KAM in comparison with current DEA models. The process of applying KAM can be seen in the following figure.

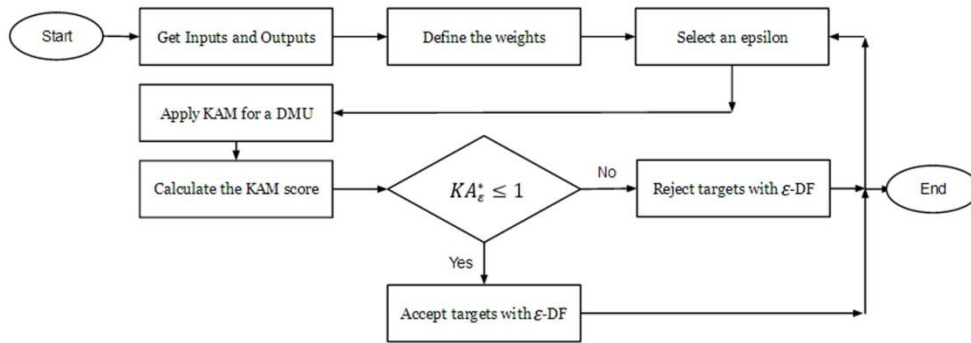


Figure: How to Apply KAM.

User may select different values of epsilon in interval $[0, 0.1]$, and graph how KAM efficiency scores of a DMU are changed on $[0, 0.1]$. In this case we have KAM function as follows: $Y_i(\varepsilon) = KA_\varepsilon^{*l}$.

If user is interest to find a minimum KAM score for an epsilon, it is suggested to extend non-linear AM (Khezrimotlagh et al. 2013c) to non-linear KAM given by:

$$\min KA_\varepsilon^{*l} = \frac{1 + \sum_{j=1}^m W_{lj}^- (E_{lj}^- - s_{lj}^-)}{1 + \sum_{k=1}^p W_{lk}^+ (s_{lk}^+ - E_{lk}^+)}$$

Subject to

$$\sum_{i=1}^n \lambda_{li} x_{ij} + s_{lj}^- = x_{lj} + \varepsilon_{lj}^-, \text{ for } j = 1, 2, \dots, m,$$

$$\sum_{i=1}^n \lambda_{li} y_{ik} - s_{lk}^+ = y_{lk} - \varepsilon_{lk}^+, \text{ for } k = 1, 2, \dots, p,$$

$$\sum_{i=1}^n \lambda_{li} = 1, \lambda_{li} \geq 0, \text{ for } i = 1, 2, \dots, n, s_{lj}^- \geq 0, \text{ for } j = 1, 2, \dots, m, s_{lk}^+ \geq 0, \text{ for } k = 1, 2, \dots, p.$$

where $W_{lj}^- = w_{lj}^- / \sum_{j=1}^m w_{lj}^- x_{lj}$, $W_{lk}^+ = w_{lk}^+ / \sum_{k=1}^p w_{lk}^+ y_{lk}$, $E_{lj}^- = \varepsilon / w_{lj}^-$ and $E_{lk}^+ = \varepsilon / w_{lk}^+$.

Theorem 1: The targets of the above non-linear KAM is always on the Farrell frontier.

Theorem 2: If $w_{lj}^- = 1/x_{lj}$, $w_{lk}^+ = 1/y_{lk}$ and $\varepsilon = (\varepsilon^-, \varepsilon^+) = 0 \in \mathbb{R}_+^{m+p}$, non-linear KAM is Slack-Based Measure (SBM) proposed by Tone (2001).

Theorem 3: Non-linear KAM can be transferred to linear form similar to SBM.

Theorem 4: The score of non-linear KAM is always less than or equal the score of linear KAM.

The above theorems clearly illustrate the significant and robustness of KAM in comparison with current DEA models. Indeed, non-linear KAM can consider any possible weights of DMUs and its scores are always in the interval [0,1]. Now KAM function can be defined as follows for $\varepsilon = (\varepsilon^-, \varepsilon^+) \in \mathbb{R}_+^{m+p}$:

$$Y_l(\varepsilon) = \frac{1 + \sum_{j=1}^m W_{lj}^-(E_{lj}^- - s_{lj}^{-*})}{1 + \sum_{k=1}^p W_{lk}^+(s_{lk}^+ - E_{lk}^{+*})}$$

User may change the values of epsilon for KAM function, and graphs it. When the epsilon components are $\varepsilon_j^- = \varepsilon x_j$, and $\varepsilon_k^+ = \varepsilon y_k$, for $\varepsilon \in \mathbb{R}_+$, $j = 1, 2, \dots, m$, and $k = 1, 2, \dots, p$, if the value of epsilon increases the score of non-linear KAM decreases which shows that KAM function is always decreasing in this case.

The results of linear KAM and non-linear KAM can be read with many different views. For instance, one view is that KAM says if a very small error is introduced in variables of a DMU, how much effects will have on its technical efficiency score. If the differences between the scores of 0-KAM and ε -KAM ($\varepsilon \neq 0$) is negligible, KAM suggests that the measured technical efficiency score is significant. Otherwise, KAM does not know that DMU as efficient. Khezrimotlagh *et al.* (2013a) also proposed the following definitions which illustrates the differences between efficiency and technical efficiency meanings:

Definition: A technical efficient DMU is KAM efficient with ε -DF in inputs and outputs if $KA_0^* - KA_\varepsilon^* \leq \delta$. Otherwise, it is inefficient with ε -DF in inputs and outputs. The proposed amount for δ is $10^{-1}\varepsilon$ or $\varepsilon/(m + p)$.

Another view is that KAM measures the instabilities of the Farrell frontier. KAM says if the Farrell frontier gets a small width (or gets thickness) are the known technically efficient DMUs best performers yet?

Moreover, since epsilon is a vector with $m + p$ dimensions, it can at least be selected 2^{m+p} times by selecting some of its components as 0. For instance, if $\varepsilon = (\varepsilon^-, \varepsilon^+) = (0, \varepsilon^+) \in \mathbb{R}_+^{m+p}$, KAM is KM and if $\varepsilon = (\varepsilon^-, \varepsilon^+) = (\varepsilon^-, 0) \in \mathbb{R}_+^{m+p}$, KAM is AM. Therefore, each of these selections of epsilon provides different views of KAM results. From these outcomes user can test each factor by selecting different components for epsilon. For instance, if epsilon is considered as $(0, 0, \dots, 0, \varepsilon x_{lm}, 0, \dots, 0)$, it says if the last input value is increased by εx_{lm} how much effect does it have in the measured efficiency score of DMU_l.

CONCLUSIONS

In this paper it is illustrated how an epsilon can be selected for KAM to measure the efficiency scores of DMUs. The KAM function and how to test each factor are also proposed.

REFERENCES

- [1] Banker, R.D., Charnes A., Cooper, W.W. (1984) Some Models for Estimating Technical and scale Inefficiencies in Data Envelopment Analysis. *Management Science* 30(9): 1078-1092.
- [2] Charnes A., Cooper W.W., Rhodes E. (1978) Measuring the efficiency of decision-making units, *European Journal of Operational Research* 2(6): 429-444.
- [3] Charnes, A., Cooper W.W., Golany, B., Seiford, L.M., Stutz, J. (1985) Foundations of data envelopment analysis and Pareto-Koopmans empirical production functions. *Journal of Econometrics* 30(1): 91-107.
- [4] Färe, R., Grosskopf, S., Kokkelenberg, E.C. (1989) Measuring Plant Capacity, Utilization and Technical Change: A Nonparametric Approach. *International Economic Rev.* 30(3): 655-666.
- [5] Khezrimotlagh, D., Salleh, S., Mohsenpour, Z. (2012a) A New Method in Data Envelopment Analysis to Find Efficient Decision Making Units and Rank Both Technical Efficient and Inefficient DMUs Together. *Applied Mathematical Sciences* 6(93): 4609-4615.
- [6] Khezrimotlagh, D., Mohsenpour, Z., Salleh, S. (2012b) Cost-Efficiency by Arash Method in DEA. *Applied Mathematical Sciences* 6(104): 5179-5184.
- [7] Khezrimotlagh, D., Salleh, S., Mohsenpour, Z. (2012c) Arash Method and Uncontrollable Data in DEA. *Applied Mathematical Sciences* 6(116): 5763-5768.
- [8] Khezrimotlagh, D., Mohsenpour, Z., Salleh, S. (2012d) Comparing Arash Model with SBM in DEA. *Applied Mathematical Sciences* 6(104): 5185-5190.
- [9] Khezrimotlagh, D., Salleh, S., Mohsenpour, Z. (2012e) Airport Efficiency with Arash Method in DEA. *Journal of Basic and Applied Scientific Research* 2(12): 12502-12507.
- [10] Khezrimotlagh, D., Salleh, S., Mohsenpour, Z. (2012f), Benchmarking Inefficient Decision Making Units in DEA, *Journal of Basic and Applied Scientific Research*, 2(12): 12056-12065.
- [11] Khezrimotlagh, D., Salleh S., Mohsenpour, Z. (2013a) A New Method for Evaluating Decision Making Units in DEA, *Journal of the Operational Research Society*, 65(1): 694–707.
- [12] Khezrimotlagh, D., Salleh, S., Mohsenpour, Z. (2013b) A New Robust Mixed Integer Valued Model in DEA, *Applied Mathematical Modelling*, 37(24): 9885-9897.
- [13] Khezrimotlagh, D., Salleh, S., Mohsenpour, Z. (2013c) Nonlinear Arash Model in DEA. *Research Journal of Applied Sciences. Engineering and Technology* 5(17): 4268-4273.
- [14] Khezrimotlagh, D., Mohsenpour, P., Salleh, S., Mohsenpour, Z. (2013d). Decision Making Units with Integer Values in Data Envelopment Analysis. In: Banker B., A. Emrouznejad, H. Bal, I. Alp, M. Ali Gengiz (2013), *Data Envelopment Analysis and Performance Measurement: Proceedings of the 11th International Conference of DEA, June 2013, Samsun, Turkey, ISBN: 9781854494771*.
- [15] Khezrimotlagh, D., Mohsenpour, P., Salleh, S., Mohsenpour, Z. (2013e). A Review on Arash Method in Data Envelopment Analysis. In: Banker B., A. Emrouznejad, H. Bal, I. Alp, M. Ali Gengiz (2013), *Data Envelopment Analysis and Performance Measurement: Proceedings of the 11th International Conference of DEA, June 2013, Samsun, Turkey, ISBN: 9781854494771*.
- [16] Khezrimotlagh, D. (2014) Profit Efficiency by Kourosh and Arash Model, *Applied Mathematical Sciences*, 8(24): 1165-1170.
- [17] Lozano, S., Gutierrez, E. (2011) Slacks-based measure of efficiency of airports with airplanes delays as undesirable outputs. *Computers and Operations Research* 38(1): 131-139.
- [18] Mohsenpour, P., Munisamy, S., Khezrimotlagh, D. (2013), Revenue Efficiency and Kourosh Method in DEA, *Applied Mathematical Sciences* 7(140): 6961-6966.
- [19] Simar, L., Wilson, P. (1998) Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models. *Management Science* 44(1): 49-61.
- [20] Tone, K. (2001) A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research* 130(3): 498-509.