한국지능정보시스템학회논문지 제7권 제1호 2001년 6월(pp. 63~79)

Evaluating Efficiency of Life Insurance Companies Utilizing DEA and Machine Learning

Han Kook Hong
Department of Mis, Dong-Eui University
(honghk@hyomin.dongeui.ac.kr)

Jae Kyeong Kim
School of Business Administration,
Kyung Hee University
(jaek@khu.ac.kr)

Data Envelopment Analysis(DEA), a non-parametric productivity analysis tool, has become an accepted approach for assessing efficiency in a wide range of fields. Despite of its extensive applications and merits, some features of DEA remain bothersome. DEA offers no guideline about to which direction relatively inefficient DMUs improve since a reference set of an inefficient DMU, several efficient DMUs, hardly provides a stepwise path for improving the efficiency of the inefficient DMU. In this paper, we aim to show that DEA can be used to evaluate the efficiency of life insurance companies while overcoming its limitation with the aids of machine learning methods.

Key words: Data envelopment analysis, Life insurance company, Self-organizing map, Machine learning

1. Introduction

In the age of globalization, autonomy, and high competition, life insurance companies want to set up their management strategies in order to improve the efficiency of operation and to gain a competition advantages. In order to do so, life insurance companies need an appropriate tool to precisely measure the efficiency of their operations. Based on the operational efficiencies, they set up their improvement strategies to reach more efficient companies. Even though there have been many studies about benchmarking, it is hard to find the study which tells detailed guidelines about how to find a target of benchmarking.

This research suggests a hybrid methodology to find an improvement path of life insurance companies. The improvement path helps a life insurance company to do an efficient operational management. Our suggested hybrid methodology is based on two techniques; one is a Tier Analysis based on Data Envelopment Analysis (DEA) and the other one is a cluster analysis based on self-organizing map (SOM).

The DEA model is a fractional linear program that aims to assess the comparative efficiency of Decision-Making Units (DMUs) where there are multiple possibly incommensurate inputs and outputs. DEA was developed by Charnes et al. [3] as a generalization of the framework of Farrell [6] on the measurement of productive efficiency. They generalized Farrell's model and allowed it to cast in the form of a fractional expression or ratio.

Numerous researches on efficiency measurement of real life problems using DEA have been conducted. DEA has been tested empirically in many settings including university departments [2], criminal superior courts [9], hospitals [11], rate-collection units [15], vehicle maintenance sections [4], and branch network of a bank [5].

As the earlier list of applications suggests, DEA can be a powerful tool by the following characteristics: First, it doesn't require an assumption of a functional form relating inputs to outputs. Second, it allows managers to consider simultaneously multiple inputs and multiple outputs of a DMU. Third, it provides managers with a procedure to differentiate efficient DMUs from the inefficient. Fourth, it pinpoints the sources and the amount of deficiency for each of the inefficient DMUs. Finally, it can be used to detect specific inefficiencies that may not be detectable through other techniques such as linear regression or ratio analyses.

Tier analysis is a kind of technique that can be used to cluster DMUs together according to their efficiency levels. SOM is one of well-known clustering tools for grouping similar DMUs according to input/output patterns, for the inefficient DMU to select one efficient DMU in a reference set as a benchmarking target. With the efficient tiers identified by the tier analysis, it can provide the guidelines for stepwise improvements of inefficient DMUs.

Cluster analysis refers a collective analysis that regards similar DMUs as a single cluster. The similarity between DMUs is determined by the domain-specific knowledge. The basic idea is the DMUs in a single cluster share some common domain-specific knowledge. So it will be easy for a less inefficient company to be more efficient if the company tries to mimic (or follow) the management strategy of a more efficient companies in the same class.

In this study, we utilize our methodology to evaluate efficiencies of 29 life insurance companies located in South Korea.

The remainder of the paper is structured as follows. The next chapter presents a review of literature on DEA and the productivity measures of life insurance companies. This is followed by a description of the research methodology. The subsequent chapter presents experimental design and results. The concluding remarks are presented in the last chapter.

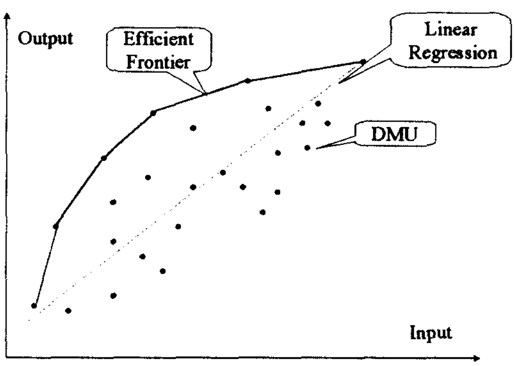
2. Literature Review

2.1 DEA

DEA is itself a basic concept that can be given a variety of forms in diverse applications. In any of these forms, however, it can be applied to empirical data in a relatively straightforward manner via different types of models to obtain estimates of the efficiency of different DMUs defined as the organizations or entities which are responsible for converting inputs into outputs.

DEA uses observed or reported values of multiple outputs and inputs for each DMU and makes repeated use of the same optimizing principle to select subsets of efficient DMUs. Each DMU is evaluated by DEA in this same manner and the sources and amounts of its inefficiencies are identified and estimated.

DEA involves an alternative principle for extracting information about a population of observations such as those shown in <Fig. 1>. In contrast to parametric approach whose object is to optimize a single regression plane (dotted line) through the data, DEA optimizes on each individual observation with an objective of calculating a discrete piecewise frontier (solid line) determined by the set of Pareto-efficient DMUs. That is, the focus of DEA is on the individual observations as represented by the *n* optimizations (one for each observation) required in DEA analysis, in contrast to the focus on the average and estimation of parameters that are associated with single-optimization statistical approaches.



<Fig. 1> Comparison of DEA with regression analysis

The parametric approach requires the imposition of a specific functional form (e.g., a regression equation) relating the independent variables to the dependent variables. The selected functional form also requires specific assumptions about the distribution of the error terms and many other restrictions such as factors earning the value of their marginal product. In contrast, DEA does not require any assumptions about the functional form. DEA calculates a maximal performance measure for each DMU relative to all other DMUs in the observed population with the sole requirement that each DMU lies on or below the extreme frontier. Each DMU not on the frontier is scaled against a convex combination of the frontier facet.

The solid line in <Fig. 1> represents a frontier derived by DEA from data on a population of DMUs, each utilizing different amounts of a single input to produce various amounts of a single output. It is important to note that DEA calculations produce only relative efficiency measures. The relative efficiency of each DMU is calculated in relation to all the other DMUs, using the actual observed values for the inputs and outputs of each DMU.

For each inefficient DMU that lies below the frontier, DEA identifies the sources and levels of inefficiencies determined by the comparison to a single referentative DMU or a convex combination of other referentative DMUs located on the efficient frontier, which utilize the same level of inputs and produce the same or a higher level of outputs. The calculated improvements for inefficient DMUs are indicative of potential improvements because the projections are based on the revealed best-practice performance of comparable DMUs located on the efficient frontier.

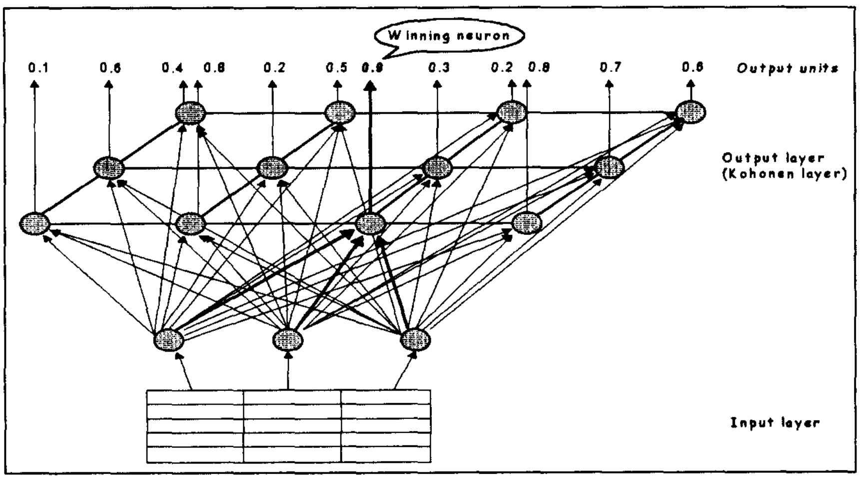
Rapid development in DEA has attracted attention from researchers in the field of Multiple Criteria Decision Making(MCDM). One of the earliest attempts to integrate MCDM procedures and DEA techniques was made by Golany [7], who suggests an interactive Multiple Objective Linear Programming(MOLP) procedure for estimating a target set of output levels given the available input levels of a DMU. Stewart [13] contrasts the concept of relative efficiency in DEA with that of Pareto optimality in MCDM and discusses some issues in applying interactive MCDM techniques for solving the weight restriction problem in DEA.

2.2 SOM

SOM uses an unsupervised learning scheme to train the neural network(see [10, 12]). Unsupervised learning is comprised of those techniques for which the resulting actions or desired outputs for the training sequences are not known. The network is only told the input vectors, and the network self-organizes these inputs into categories.

Each link between nodes in the input layer and nodes in the output layer has an associated weight. The net input into each node in the output layer is equal to the weighted sum of the inputs. Learning proceeds by modifying these weights from an assumed initial distribution with the presentation of each input pattern vector. This process identifies groups of nodes in the output layer that are close to each other and respond in a similar manner. A particular group of units together forms an output cluster. The topology preserving mappings from the inputs to the clusters reflect the existing similarities in the inputs and capture any regularities and statistical features, and model the probability distributions which are present in the input data.

SOM uses competitive learning. When an



<Fig. 2> Winning neuron and its path

input pattern is imposed on the network, one output node is selected from among all the output nodes as having the smallest Euclidean distance between the presented input pattern vector and its weight vector. This output unit is declared the winner in the competition among all the neurons in the output layer. Only the winning neuron generates an output signal from the output layer. All the other neurons have a zero output signal (see <Fig. 2>).

The input and weight vectors are usually normalized in a SOM so that they have values between 0 and 1 inclusive. If the dot products between the normalized input vector $\hat{X}$ and a normalized set of weight vectors $\hat{W}_j$ are determined, the neuron with the largest dot product (the one with the smallest Euclidean distance) is declared to be the winner. Thus the winner is the vector obtained from the expression:

$$\max_j \left(\hat{X}^t \ \hat{W}_j \right)$$

As learning involves adjustment of weight vectors, learning with this particular input pattern is restricted to lateral interconnections with immediately neighboring units of the winning neuron in the output layer. Adjusting their weights closer to the input vector carries out learning for the nodes within the neighborhood. The size of the neighborhood is initially chosen to be enough large to include all units in the output layer. However, as learning proceeds, the size of the neighborhood is progressively reduced to a pre-defined limit. Thus during these stages, fewer neurons have their weights adjusted closer to the input vector. Lateral inhibition of weight vectors that are distant from a particular input pattern may also be carried out.

2.3 Efficiency evaluation of Life Insurance Company

A method of analyzing productivity of a life insurance company is, in general, to represent the relationship of inputs and outputs as a generalized Leontief profit function and to estimate parameters of the function (see [16]). However, the life insurance industry has such an uncertain management environment as inaccuracy of price information on inputs and outputs, unbalance of the amount of inputs and outputs due to monopoly or duo-poly, the exit from or entry into the industry, and government regulations on insurance rate. These limitations prevent the parametric method from being used, which needs strict assumptions on a population.

DEA, as a non-parametric method, overcomes the shortcomings of the parametric method. Because it evaluates relative efficiency of inputs and outputs and doesn't need to further consider technical relationships between inputs and outputs.

But, little research has been made to measure the efficiency of life insurance companies using DEA. The real difficulty of those efficiency related studies lies in measuring the production of the insurance industry. Several authors have emphasized this problem. As Hornstein and Prescott [8] explain "there is not even a conceptual definition of the output to guide the construction of a reasonable measure of its product. Without a conceptual measure, it is not clear what data should be collected and how they should be used to compute an output measure".

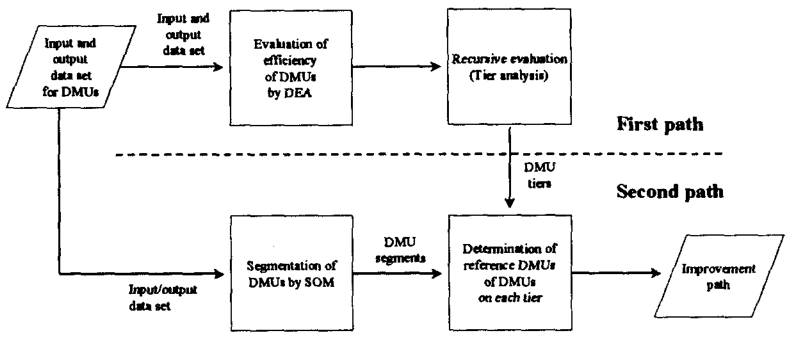
Two alternatives are often suggested: on one hand, premiums or incurred losses, and on the other hand, the number of contract policies. In recent papers, premiums earned, losses and financial investments are used as a proxy for nominal output. The pitfalls of this approach are well illustrated by Hornstein and Prescott [8].

3. A Methododlogy

In this chapter, our research framework is presented as shown in <Fig. 3>. It is comprised of two parallel paths from the starting point of analysis. In the first path, we evaluate the efficiency of DMUs through DEA. And we repeatedly evaluate DMUs, which are classified as inefficient. We call this process "Tier analysis". In the second path, the same set of DMUs is clustered into a number of segments via SOM, which is one of clustering tools. With these segments of DMUs by the SOM and the DMU tiers by DEA, a set of reference DMUs for DMUs on each tier are determined. We call this set of reference DMUs "an improvement path", which inefficient DMUs can track for improving their efficiencies.

3.1 Definition of input and output data set for DMUs

In this paper, we propose an evaluation model, with four inputs and two outputs, of life insurance companies in Korea as shown in <Fig. 4>. The basic inputs (resources) used by each DMU are net operating expenses(NOE), the number of office workers(NOW), the number of sales persons(NSP), and the number of branch offices(NBO). The net operating expenses can be calculated by subtracting income expenses from provision expenses, such as labor wages, general administration expenses, welfare expenses, and salesman recruiting expenses. The factors such as the number of office workers, the number of sales persons, and the number of branch offices are also included as inputs in this analysis, because a life insurance company belongs to a labor-intensive industry and they can be used as indexes for representing labor efficiency.

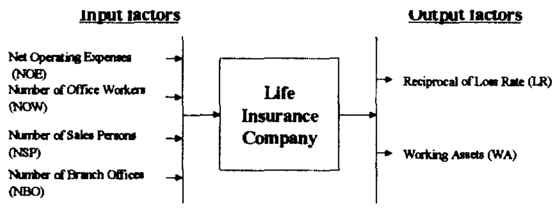


<Fig. 3> Framework of analysis

The outputs for DMUs include the reciprocal of loss rate(LR) and the working assets(WA) respectively. Among these output factors, the former can be determined as the ratio of premium receipts to claims paid. Because the ratio considers both the premium receipts and the claims paid at the same time, it, as a relative productivity index, can reveal the efficiency of a life insurance company more clearly than any other types of factors. The other output factor, the working assets, is comprised of cash, deposits, trust, securities, and real estate, which are the sources of property investment. Because a life insurance company generally makes profit through the business of insurance and finance, the amount of the working assets can play an important role in finance business. So it has to be included within this study.

On the other hand, we do not include the number of insurance contracts as a factor for evaluating DMUs. That is because domestic life insurance companies sell several types of life insurance products and their prices are different among them. Considering the number of insurance contracts as a factor could introduce an uncertain measurement.

DMUs used in this analysis are 29 life insurance companies in South Korea. Though they have a little difference in their operational activities but the Korean government has regulated their activities altogether, and they have carried out almost similar business activities. Therefore it seems plausible that we claim that we can compare their relative performance productivity.



<Fig. 4> Evaluation model of life insurance company

<Table 1> summarizes the input and output variables used in this analysis.

<Table 1> Summary of input and output variables

Variable		Measurement
Input factors	Net Operating Expenses (NOE)	By subtracting income expenses from provision expenses, such as labor wages, general administration expenses, welfare expenses, and salesman recruiting expenses
	Number of Office Workers (NOW)	The number of persons who manage sales persons and staffs of the head office
	Number of Sales Persons (NSP)	The number of persons who directly do a business with customers
	Number of Branch Offices (NBO)	The total number of branch offices geographically dispersed
Output factors	Reciprocal of Loss Rates (LR)	The ratio of Premium Receipts to Claims Paid
	Working Assets (WA)	It is comprised of cash, deposits, trust, securities, and real estate, which are the sources of property investment.

3.2 Evaluation of efficiency of DMUs using DEA

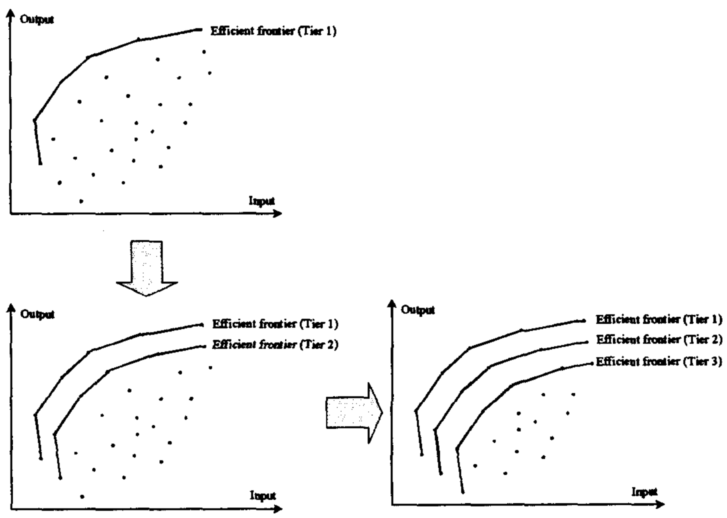
We used DEA to evaluate the efficiencies of DMUs. DEA determines the most productive group of the DMUs and the group of less-productive DMUs. That is, the DMUs are clustered into an efficient group or an inefficient one by DEA.

3.3 Recursive evaluation (Tier Analysis)

A similar approach to clustering DMUs by DEA was presented by Thanassoulis [14]. However, the clusters on that study were not made by their efficiency levels but by the characteristics of the input resource mix. Tier analysis that we propose is a kind of technique that can be used to cluster DMUs together according to their efficiency levels.

In the first step of tier analysis, we obtain the efficiency scores of the set of entire DMUs. The result of the first step should reveal the most efficient group of DMUs by indicating their scores are equal to 1.0. We call this group "Tier 1". In the second step, we proceed DEA again only with the inefficient DMUs which are not part of Tier 1. DMUs whose efficiency scores in the second step are equal to 1.0 are Tier 2. The same procedure can be repeated during the number of remaining inefficient DMUs is at least three times multiple of that of inputs along with outputs (4 + 2 = 6), as Banker et al. [1] have proposed, which makes it possible to appropriately discriminate efficient DMUs from inefficient ones. We call this procedure "the tier analysis" because DMUs that belong to each tier form the efficient production frontier in each step as shown in <Fig. 5>.

<Fig. 5> shows that DMUs on the tier 1 are superior to those in the tier 2 and DMUs on



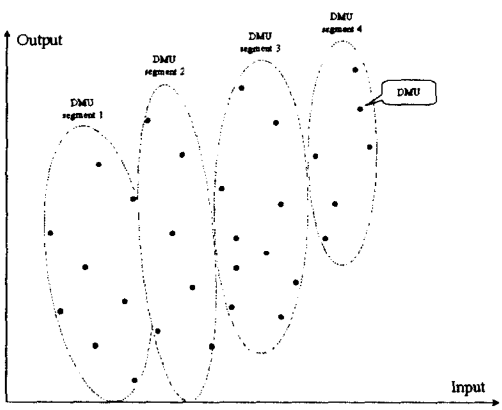
<Fig. 5> The procedure of tier analysis

tier 2 are superior to those in the tier 3. We use these DMU tiers in the second path of our analysis to determine a stepwise improvement path for each of inefficient DMUs.

3.4 Segmentation of DMUs using SOM

In the second path of our analysis, we plan to newly use a self-organizing map (SOM), which is one of clustering tools, with the DMU tiers to further suggest an improvement path for each inefficient life insurance company.

DEA offers no guidelines about to which direction relatively inefficient DMUs improve since a reference set of an inefficient DMU consists of several efficient DMUs. Hence, we make SOM group similar DMUs with the similar characteristics of the inputs, for the inefficient DMU to select an efficient DMU in a reference set as a benchmarking target (refer to <Fig. 6>).

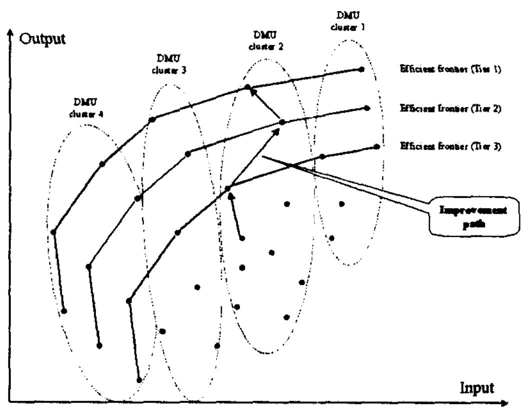


<Fig. 6> DMU segments

3.5 Determination of reference DMUs of DMUs on each tier

Efficient DMUs in the upper tier become a reference set of inefficient DMUs in the lower tier. How can we select a target reference DMU among DMUs in the reference set ? We use the SOM in advance to find a target reference DMU in the upper tier, which has the similar input characteristics with inefficient DMUs that lie on the lower tier. Refer to <Fig. 7>.

Once the tiers by DEA and the DMU segments by SOM have been identified, we determine a stepwise path for improving the efficiency of each inefficient DMU as shown in <Fig. 7>.



<Fig. 7> The reference set of the inefficient DMUs on each tier

4. A Case Study : Life Insurance Companies

4.1 Evaluation of the efficiency of the DMUs using DEA

DMUs that we used are 29 life insurance companies. We use the Charnes-Cooper-Rhodes (CCR) ratio model of DEA to evaluate the efficiency of them. Results of DEA indicate the relatively efficient life insurance companies and the less efficient ones (refer to <Table 2>).

<Table 2> Life insurance companies' efficiency ratings

Life insurance companies (DMUs)	Efficiency Rating	Reference set
C1	0.62	C3, C5, C17, C25
C2	0.74	C3, C5, C17
C3	1.00	
C4	0.80	C5, C25
C5	1.00	
C6	0.39	C3, C5, C17
C7	0.60	C3, C5, C17, C25
C8	0.42	C3, C5, C17
C9	0.49	C3, C5, C25
C10	0.63	C3, C5, C17
C11	0.75	C5, C25
C12	0.83	C3, C5, C17
C13	0.84	C3, C17
C14	0.42	C3, C5, C17, C25
C15	0.54	C3, C5, C17, C25
C16	0.7	C5, C25
C17	1.00	
C18	0.40	C5, C17, C25
C19	0.81	C5, C25
C20	0.45	C3, C5, C17
C21	0.54	C5, C25
C22	0.74	C5, C25
C23	0.35	C5, C25
C24	0.68	C3, C5, C17
C25	1.00	
C26	0.32	C3, C17, C25
C27	0.76	C3, C17, C25
C28	0.44	C3, C17, C25
C29	0.80	C5, C25

The table shows that four life insurance companies, C3, C5, C17 and C25 are best-practiced companies with DEA productivity rating of 100 percent. Company C1 is less productive with DEA productivity rating of 62 percent, suggesting that it could provide its current mix and volume of outputs with only about 62 percent of the resources it actually consumes. Company C2 has DEA productivity rating of 74 percent indicating that it is using about 26 percent excess resources. In fact, 25 of the 29 companies are using excess resources. These findings indicate that there is room that the 25 life insurance companies could make substantial productivity improvements and cost reductions.

4.2 Clustering the DMUs through the Tier Analysis

We group 29 life insurance companies together into four tiers by the tier analysis. The efficiency score itself is not important in this time. Only what matters is to which tier each company belongs.

1) In the first tier analysis, the efficient DMUs by DEA form the "tier 1" and the remaining inefficient DMUs become the candidates for the second application of DEA. The result of the first tier analysis are summarized as <Table 3>. Note that, in the column "DMUs" of the format C_nN, n indicates the tier on which each company is and N indicates the number of DMU.

<Table 3> Clustering of life insurance companies by the tier analysis - tier 1

Group(Tier)	DMUs	Reference set
1	$C_1$3 $C_1$5 $C_1$17 $C_1$25	No reference set

2) After the first tier analysis, we proceed DEA again only with the inefficient DMUs which are not part of Tier 1. DMUs whose efficiency scores are 1.0 in the second tier analysis are Tier 2 (refer to <Table 4>). The same procedure should be repeated during the number of remaining inefficient DMUs is at least three times multiple of that of inputs along with outputs.

<Table 4> Clustering of life insurance companies by the tier analysis - tier 2

Group(Tier)	DMUs	Reference set
2	$C_2$1	$C_1$3 $C_1$5 $C_1$17 $C_1$25
	$C_2$12	$C_1$3 $C_1$5 $C_1$17
	$C_2$13	$C_1$3 $C_1$17
	$C_2$27	$C_1$3 $C_1$17 $C_1$25
	$C_2$28	$C_1$3 $C_1$17 $C_1$25
	$C_2$29	$C_1$5 $C_1$25

3) After the third tier analysis, the results can be summarized as <Table 5> and <Table 6>. In our application of 29 life insurance companies, the fourth tier is the last one in the tier analysis.

<Table 5> Clustering of life insurance companies by the tier analysis - tier 3

Group(Tier)	DMUs	Reference set
3	$C_3$2	$C_2$12
	$C_3$4	$C_2$12
	$C_3$7	$C_2$1 $C_2$12
	$C_3$15	$C_2$12 $C_2$29
	$C_3$16	$C_2$12 $C_2$29
	$C_3$18	$C_2$12 $C_2$27 $C_2$29
	$C_3$19	$C_2$12 $C_2$29
	$C_3$22	$C_2$12 $C_2$29
	$C_3$24	$C_2$12 $C_2$13 $C_2$27
	$C_3$26	$C_2$12 $C_2$27 $C_2$29

<Table 6> Clustering of life insurance companies by the tier analysis - tier 4

Group(Tier)	DMUs	Reference set
4	$C_4$6	$C_3$2 $C_3$16
	$C_4$8	$C_3$2 $C_3$16 $C_3$24
	$C_4$9	$C_3$7
	$C_4$10	$C_3$4 $C_3$7 $C_3$24
	$C_4$11	$C_3$4 $C_3$18 $C_3$22 $C_3$26
	$C_4$14	$C_3$16
	$C_4$20	$C_3$2 $C_3$4 $C_3$18 $C_3$24
	$C_4$21	$C_3$16 $C_3$22 $C_3$26
	$C_4$23	$C_3$2 $C_3$4 $C_3$16 $C_3$24

4.3 Clustering the DMUs using SOM

<Fig. 8> shows the results of clustering 29 DMUs by SOM. Here come four clusters out. The number in each cluster indicates each company, which belongs to that cluster. Therefore, cluster 1 has one member, such as C3, and cluster 2 has 7 DMUs in it.

(1)	(2)
C3	C2, C4, C6, C8 C9, C12, C14
(3)	(4)
C1, C5	C7, C10, C11, C13 C15, C16, C17, C18 C19, C20, C21, C22 C23, C24, C25, C26 C27, C28, C29

<Fig. 8> The clustering result of SOM

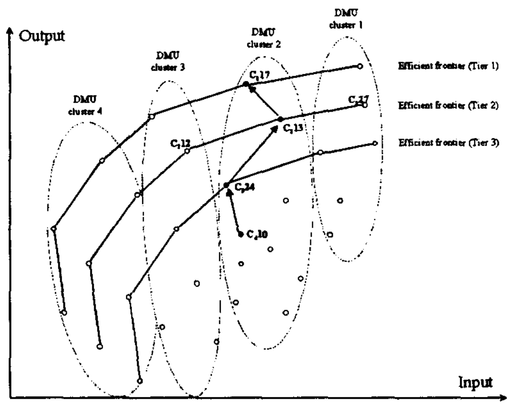
<Table 7> summarizes the characteristic of each cluster in detail. It shows four each cluster, the number of DMUs which belong to each cluster, and average values of inputs and outputs.

4.4 Determining the reference DMU of the inefficient DMU on each tier

By the tier analysis, 29 life insurance companies are divided into four different tiers according to their efficiency level. And by SOM, DMUs on the lower tiers can find the way for improving efficiencies of them. How can it be done? DMUs on each tier can improve their efficiencies through finding only one reference DMU on the very upper tier, which shares the similar characteristics with them.

For example, C_324 on tier 3 has a reference set that consists of efficient DMUs, such as C_212, C_213, and C_227, on the upper efficient frontier 2 (tier 2). Among them, we choose C_213 as a benchmarking target, since it belongs to the same segment with C_324 by SOM.

Based on the previous results that identify a benchmarking reference of each DMU on each tier, we can finally determine the stepwise improvement path for each DMU on each tier except tier 1. For example, looking at the <Fig. 9>, we are able to determine a path like C_410 -> C_324 -> C_213 -> C_117 as an improvement path for C_410.



<Fig. 9> Improvement path for a DMU C410 on tier 4

<Table 7> Characteristics of each cluster

Cluster	Count	NOE (Avg)	NOW (Avg)	NSP (Avg)	NBO (Avg)	LR (Avg)	WA (Avg)
1	1	1,411,258	7,912	58,415	1,711	1.07	33,016,684
2	7	113,017	1,554	8,197	374	0.762	1,769,286
3	2	758,874	6,769	53,760	1,604	1.02	16,180,759
4	19	28,436	396	1,656	82	1.35	300,244

<Table 8> Input/output factors of DMUs on an improvement path

Group (Tier)	DMU (Company)	Input Factors				Output Factors	
		NOE	NOW	NSP	NBO	LR	WA
1	$C_1 17$	26,909	**259**	1,723	98	**0.686**	339,621
2	$C_2 13$	26,521	401	1,892	101	0.450	**349,576**
3	$C_3 24$	**25,761**	389	1,672	**97**	0.449	251,910
4	$C_4 10$	32,265	394	1,571	131	0.456	253,137

As shown in <Table 8>, $C_3 24$ as the first benchmarking DMU on the improvement path toward $C_1 17$ consumes less net operating expenses(NOE) and less number of branch offices(NBO) than those of $C_4 10$, whereas a level of outputs is similar. Although $C_2 13$, as the second benchmarking DMU, has a similar level of input factors with$C_3 24$, the level of working assets(WA) is much higher than that of $C_3 24$. At last, $C_1 17$ generally spends less inputs, especially the number of office workers (NOW), than $C_2 13$, but it generates much more output, i.e., the reciprocal of loss rate.

5. Conclusions

DEA is good at estimating the "relative" efficiency of a DMU, as it can tell us how well we are doing compared to our peers. But, because DMUs are directly compared against a peer or combination of peers, DEA offers no guideline about to which direction relatively inefficient DMUs improve. Also it doesn't provide the stepwise path for improving the efficiency of each inefficient DMU. In order to overcome this limitation of DEA, we suggest a hybrid methodology utilizing the machine learning and DEA.

It is comprised of two parallel paths from the starting point of analysis. In the first path, we evaluate the efficiency of DMUs through DEA. After that, we clustered the DMUs together through the tier analysis, which recursively apply the DEA to the remaining inefficient DMUs. In the other path, the same set of DMUs is clustered into some segments via the SOM, which is one of clustering tools. With these segments of DMUs by the SOM and the DMU tiers by DEA, a set of reference DMUs for DMUs on each tier are determined. We call this set of reference DMUs "Improvement path", which inefficient DMUs can track for improvement of their efficiencies.

In conventional DEA, it only (1) identifies inefficiencies, (2) identifies comparable efficient units, (3) locates slack resources. But, we provide more information about targets for inefficient DMUs and about stepwise improvement path.

We resolved the limitations of the DEA that are listed in chapter 1. First, the conventional DEA provides a reference set (multiple efficient DMUs) for each inefficient DMU. It cannot give a hint on which direction relatively inefficient DMUs improve. But, since we utilize SOM as a tool for clustering DMUs according to the similarity of inputs, we can choose one DMU on reference set as a benchmarking target for each inefficient DMU.

Second, the conventional DEA cannot provide information about a continuous improvement path. It simply gives us information about the identification of inefficient DMUs and slack variables via reference set. We can resolve this problem and provide the information about continuous improvement path through using the DMUs clusters by SOM and a reference company by tier analysis.

However, the present research has a number of limitations. They can be also the topics for further researches: selection of input and output variables, evaluation of appropriateness of non-parametric methods, and inclusion of qualitative factors in output variables. Among them, selecting appropriate input and output variables is one of main issues to study further. That is because, in a DEA model, as the number of input and output variables and the number of DMUs grow, the efficiency of the model can be monotonically increased. By the way, analyzing the relationship between addition or deletion of input variables, output variables and DMUs, and the efficiency of the model must be preceded to make a first decision on selecting input and output variables.

Current practice on management evaluation of life insurance companies in South Korea has focused on their capability of growth, productivity, profitability, soundness and publicity. Therefore an extended DEA model including qualitative as well as quantitative data is needed to measure the efficiencies of DMUs more accurately. Such a problem can also make up the research about DEA methodologies.

References

[1] Banker, R. D. and Morey, R. C., "The use of Categorical Variables in Data Envelopment Analysis", Management Science, 32(1986), 1613-1627.

[2] Beasley JE. "Comparing university departments", OMEGA, 18/2(1990), 171-183.

[3] Charnes A, Cooper WW, Rhodes E. "Measuring the efficiency of decision making units", European Journal of Operational Research, 2(1978), 429-444.

[4] Clark RL. "Evaluating USAF vehicle maintenance productivity over time: An application of data envelopment analysis", Decision Sciences, 24(1992), 376-384.

[5] Drake L, Howcroft B. "Relative efficiency in the branch network of a UK bank: An empirical study", OMEGA, 22/1(1994), 83-90.

[6] Farrell MJ. "The measurement of productive efficiency", Journal of the Royal Statistical Society. Series A, 120/3(1957), 253-290.

[7] Golany, B., "An interactive MOLP procedure for the extension of DEA to effectiveness analysis", Journal of the Operational Research Society, 39/8(1988), 725-734.

[8] Hornstein A, Prescott EC. "Measuring of the insurance sector output", The Geneva Papers on Risk and Insurance, 59(1991), 191-206.

[9] Lewin AY, Morey R, Cook T. "Evaluating the administrative efficiency of courts", OMEGA, 10(1982), 404-411.

[10] Michael JA, Berry Gordon Linoff. "Data Mining Techniques for Marketing, Sales and Customer Support", John Wiley & Sons Inc. 1997.

[11] Rutledge R, Parsons S, Knaebel R. "Assessing hospital efficiency over time: An empirical application of data envelopment analysis", Journal of Information Technology Management, 6/1(1995), 13-23.

[12] Sabrina Sestito, Tharam S Dillon. "Automated Knowledge Acquisition", Prentice Hall. 1994.

[13] Stewart, T.J., "Relationships between data envelopment analysis and multiple criteria decision making", Journal of the Operational Research Society, 47/5(1996), 654-665.

[14] Thanassoulis E. "A Data Envelopment Analysis Approach to Clustering Operating Units for Resource Allocation Purposes", Omega, Int. J. Mgmt Sci, 24/4(1996), 463-476.

[15] Thanassoulis E, Dyson Rg, Foster MJ. "Relative efficiency assessments using data envelopment analysis: An application to data on rates departments", Journal of the Operational Research Society; 38/5(1987), 397-411.

[16] Weiss MA. "Efficiency in the Property Liability Insurance Industry", Journal of Risk and Insurance, 58(1991), 452-479.

국문요약

자료봉합분석과 기계학습을 이용한 생명보험회사의 효율성 평가

홍한국*
김재경**

비모수적인 효율성 분석기법인 자료봉합분석(Data Envelopment Analysis) 은 현재 학교, 은행, 병원 등 여러 분야에서 각 조직의 효율성을 평가하는데 응용되고 있다. 그러나, 방법론 적용에 있어 상대적으로 비효율적인 의사결정단위의 참조집합이 1개 이상의 효율적인 의사결정단위로 구성되어 있어, 어느 방향으로 개선해야 할지 가이드라인을 제공하지 못하고 또한 효율성 크기에 따라, 비효율적인 의사결정단위의 단계적인 개선 방향을 제공하지 못한다는 불편한 점이 있다.

따라서, 본 연구에서는 이와 같은 불편한 점을 보완하기 위한 자료봉합분석과 기계학습을 이용한 하이브리드 방법론을 개발하고, 이를 국내 29개 생명보험사에 적용하여 타당성을 검증하였다.

* 동의대학교 경영정보학과 전임강사
** 경희대학교 경영학과 부교수