A Note on "The Use of Categorical Variables in Data Envelopment Analysis"
Author(s): Wagner A. Kamakura
Source: *Management Science*, Vol. 34, No. 10 (Oct., 1988), pp. 1273-1276
Published by: INFORMS
Stable URL: http://www.jstor.org/stable/2632166
Accessed: 14/11/2008 13:07

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=informs.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We work with the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact support@jstor.org.

INFORMS is collaborating with JSTOR to digitize, preserve and extend access to *Management Science*.

http://www.jstor.org

of an equilibrium, as described by equilibrium programming developed by Garcia and myself (1981). Succinctly, each person optimizes a subproblem given that the other people have optimized their own subproblems. There is no overall optimum but rather numerous subproblems each being optimized. The result is an equilibrium programming problem, which is quite different from an optimization problem.

An actual production situation is extremely complex with numerous highly intricate and sometimes unexpected activities occurring. Equilibrium programming imparts a more accurate view of this and is the basis of my example. Since the Gerchak example assumes optimization, it is less realistic.

In sum both my example and the Gerchak example are correct, but rest on different beliefs about reality and its complexities. What does seem to be clear, as I pointed out in my example, is that the "obvious" truth is not necessarily true. Reality is subtle and enigmatic, and the field of Management Science should proceed forward with models of ever keener accuracy to comprehend it.

References

GARCIA, C. B. AND W. I. ZANGWILL, *Pathways to Solutions, Fixed Points, and Equilibria*, Prentice-Hall, Englewood Cliffs, NJ, 1981.

ZANGWILL, W. I., "From EOQ Towards ZI", *Management Sci.*, 33, 10 (October 1987), 1209–1223.

MANAGEMENT SCIENCE
Vol. 34, No. 10, October 1988
Printed in U.S.A.

NOTES

A NOTE ON "THE USE OF CATEGORICAL VARIABLES IN DATA ENVELOPMENT ANALYSIS"*

WAGNER A. KAMAKURA
Owen Graduate School of Management, Vanderbilt University, Nashville, Tennessee 37203

A variant of Banker and Morey's (1986) DEA model for controllable ordinal outputs. As opposed to the original model, this version allows the comparison of a Decision Making Unit (DMU) to other DMU's operating at equal or higher levels of the ordinal outputs.
(DATA ENVELOPMENT; ANALYSIS; EFFICIENCY ANALYSIS)

In a recent paper, Banker and Morey (1986) propose an important extension of Data Envelopment Analysis (DEA) that incorporates discrete ordinal variables into the basic BCC (Banker, Charnes, and Cooper 1984) model.

This extension is discussed by Banker and Morey (1986) within two contexts. In the first, the categorical variables represent resources not under control by the DMU's. It is shown that, in this particular case, the extended DEA model is such that a DMU is only evaluated in comparison to units in the same category (formed by the different levels of the categorical variables).

* All Notes are refereed.

Accepted by Arie Y. Lewin, former Departmental Editor; received April 30, 1987. This Note has been with the author 2 weeks for 1 revision.

0025-1909/88/3410/1273$01.25
Copyright © 1988, The Institute of Management Sciences

In the second context, the discrete variables represent controllable outputs. In this note, we show that the mixed-integer LP model specified for the second context may lead to incorrect evaluations. This is illustrated with a simple numerical example. A change in the model is then proposed to alleviate the problem.

Following the notation of the original article, a categorical output in L levels is coded as $L-1$ "dummy" variables $W_{l,j}$ so that the lowest level of this output for any DMU j would be represented by $W_{l,j} = 0$ ($l = 1, 2, \cdots L-1$), the second lowest level by $W_j = (1, 0, 0, \ldots, 0)$ and the highest level by $W_{l,j} = 1$ ($l = 1, 2, \ldots, L-1$). The DEA problem is then formulated as a mixed-integer LP model (Banker and Morey 1986, pp. 1623–1624):

$$\max \sum_{l=1}^{L-1} t_l \tag{1}$$

$$\text{subject to} \quad \sum_{j=1}^{N} \lambda_j x_{ij} \le x_{i,j0}, \qquad i = 1, 2, \ldots, M, \tag{2}$$

$$\sum_{j=1}^{N} \lambda_j y_{rj} \ge y_{r,j0}, \qquad r = 1, 2, \ldots, K, \tag{3}$$

$$\sum_{j=1}^{N} \lambda_j W_{lj} - t_l = W_{lj0}, \qquad l = 1, 2, \ldots, L-1, \tag{4}$$

$$\sum_{j=1}^{N} \lambda_j = 1, \qquad \lambda_j \ge 0, \qquad j = 1, 2, \ldots, N, \tag{5}$$

$$t_{l-1} - t_l \ge 0, \qquad l = 2, \ldots, L-1, \tag{6}$$

where t_l are 0–1 integers.

The last set of constraints was included "to insure that improvements on the service orientation [the categorical output] are identified sequentially; improvements in service orientation $(\cdots)$ can only occur by sequentially converting the elements of the service vector [w's] to 1's" (Banker and Morey 1986, pg. 1624). The authors' "sole intention [with this model] is to identify the maximum gain possible in the categorical [output]."

The application of the BM model to a simple numerical example clearly indicates that it may lead to an incorrect evaluation of the DMU's. In this example we consider 5 DMU's operating with one input and producing one output at three quality levels:

DMU	Input	Output Quantity	Output Quality	0–1 Vectors w_1	w_2
A	1	2	high	1	1
B	5	4	high	1	1
C	2	2	low	0	0
D	4	3	medium	1	0
E	1	1	medium	1	0

It is quite evident, in this example, that DMU D is inefficient, since a 25–75% combination of DMU's "A" and "B" would yield a higher amount of the output at a higher quality level, while using the same amount of the input. Conversely, a 50–50% combination of the same DMU's will need less inputs to yield the same output at a higher quality.

However, application of the BM model to DMU "D" yields $\lambda_D = 1$ and all other variables equal to zero, indicating that this DMU is at the efficiency frontier!

This discrepancy is caused by the set of constraints in (6). In reality, these inequalities on the binary slacks t_l are such that if any binary descriptor W_l is at the high level, no improvement can be made on the subsequent descriptors W_i ($i = l + 1, \ldots, L - 1$). In other words, a DMU operating at the medium level could not improve output quality further, regardless of the characteristics of the other units. This happens because t_1 has to be zero (since the first descriptor is already 1) and t_2 is constrained to be no larger than t_1. On the other hand, a DMU (such as "C") operating at the lowest quality could be raised to the highest level (since $t_1 = 1$ and $t_2 = 1$ would satisfy the last set of constraints)!

This problem can be solved by realizing that constraint (6) should pertain to the level of the categorical outputs, rather than the binary slacks. Hence, one can replace (6) by

$$t_l - t_{l-1} \leq W_{l-1,j0} - W_{l,j0}. \tag{6a}$$

This new constraint assures that the 0–1 descriptors are consistent with the ordered categories, *after improvements* in the discrete output.

With this change in the constraints, the model will identify possible improvements in the categorical output, as intended. However, the model will not be capable of identifying inefficiencies which result from the continuous input-output variables. For example, when applied to DMU E, the model with the new constraints will identify the potential gain in the output quality from medium to high (compared to the efficient DMU A), but not the potential gain of one unit of the quantitative output. This can be achieved by redefining the DEA model in BM's Appendix B (p. 1626) as:

$$\max \Phi_0 + \epsilon\Big(\sum_{r=i}^{K} \bar{S}_r + \sum_{I=1}^{M} \overset{+}{S}_i + \sum_{l=1}^{L} t_l\Big)$$

$$\text{s.t.} \quad \sum_{j=1}^{N} \lambda_j Y_{rj} - \bar{S}_r - \Phi_0 Y_{rj0} = 0 \qquad (r = 1, 2, \cdots K), \tag{7}$$

$$\sum_{j=1}^{N} \lambda_j X_{ij} + \overset{+}{S}_i = X_{ij0} \qquad (i = 1, 2, \cdots M), \tag{8}$$

$$\sum_{j=1}^{N} \lambda_j W_{lj} - t_l = W_{l,j0} \qquad (l = 1, 2, \cdots L - 1), \tag{9}$$

$$\sum_{j=1}^{N} \lambda_j = 1, \tag{10}$$

$$t_l - t_{l-1} \leq w_{l-1,j0} - w_{l,j0} \qquad (l = 2, \cdots L), \tag{11}$$

$$\bar{S}_r,\ \overset{+}{S}_i,\ \lambda_j,\ \Phi_0 \geq 0. \tag{12}$$

Application of the revised model to DMU "D" in our numerical example yields $\Phi_0 = 1.167$, $t_2 = 1$, $\lambda_1 = 0.25$ and $\lambda_2 = 0.75$, indicating that a composite DMU (combining "A" and "B") would yield 16.7% more of the output at a higher output grade while using the same inputs. The results of DMU E ($\Phi_0 = 2$, $t_2 = 1$ and $\lambda_1 = 1$) indicate that output can be increased by 100% at a higher quality level while maintaining the same inputs.

While overcoming some of the limitations of the original model, the revisions suggested here have a limitation of their own; the revised model does not allow for "virtual" DMU's that combine different levels of the categorical output. For example, if we replace DMU B by B' with the same input and outputs but "medium" quality in our previous example, the revised model will conclude that DMU C cannot be made efficient by output augmentation, but must reduce input by one unit and improve output quality from "low" to "high", in comparison to DMU A. However, one can easily see that a (75, 25%)

combination of DMU's A and B' will utilize the same input and produce 25% more output at a better (though undefined) quality than C. These last results can be still obtained from the revised model, in its linear solution ($\Phi_0 = 1.25$, $\lambda_1 = 0.75$, $\lambda_2 = 0.25$, $t_1 = 1$, $t_2 = 0.75$). On the other hand, the interpretation of the improvement in the categorical output will lack meaning, since the model prescribes a change to an undefined quality level between "medium" and "high".

References

BANKER, R. D. AND R. C. MOREY, "The Use of Categorical Variables in Data Envelopment Analysis," *Management Sci.*, 32, 12 (December 1986), 1613–1627.

———, A. CHARNES AND W. W. COOPER, "Models for Estimation of Technical and Scale Inefficiencies in Data Envelopment Analysis," Management Sci., 30, 9 (September 1984), 1078–1092.