THE USE OF DATA ENVELOPMENT ANALYSIS IN AGRICULTURE IN THE PRESENCE OF INTERVAL DATA

Eliane Gonçalves Gomes *
João Alfredo de Carvalho Mangabeira
Embrapa Satellite Monitoring
Av. Dr. Júlio Soares de Arruda 803, Parque São Quirino, 13088-300, Campinas, SP, Brazil
eliane@cnpm.embrapa.br; manga@cnpm.embrapa.br
João Carlos C. B. Soares de Mello
Department of Production Engineering – Universidade Federal Fluminense
Rua Passo da Pátria 156, São Domingos, 24210-240, Niterói, RJ, Brazil
jcsmello@producao.uff.br
Lidia Angulo Meza
Department of Science and Technology – Universidade Veiga de Almeida
Rua Ibituruna 108, 4º andar, Maracanã, 20271-020, Rio de Janeiro, RJ, Brazil
lidia_a_meza@yahoo.com

This paper uses Data Envelopment Analysis (DEA) models to measure the efficiency of some agricultural producers from the Holambra district (São Paulo - Brazil). The total area of each property is one of the variables of the model, presenting two possible values: the value declared by the producer and the value measured using IKONOS II satellite images. This variable presents uncertainty in its measurement and can be represented as interval data. The efficiency frontier is constructed considering the limits of uncertainty (interval limits), that is, the smallest and greatest possible values to be assumed for the imprecise variable. In this way, a region is constructed in relation to which the DMUs have a certain membership degree. The theoretical model, called interval DEA model, is presented, as well as its application to the case of the evaluation of the efficiency of the agricultural producers of Holambra.

**** corresponding author tel.: + 55 19 3256-6030; fax: + 55 19 3254-1100***

1. INTRODUCTION

The use of Data Envelopment Analysis (DEA) models in agriculture can support agricultural decisions (and assistance organisations), by indicating the sources of inefficiency and the units, which can serve as examples of the practices adopted (benchmarks).

This paper makes use of DEA to measure the efficiency of a sample of 71 agricultural producers from Holambra, São Paulo State, Brazil. The variables used are: total area, manpower, and use of machinery as input and net revenue as output. However, the total area presents two possible values: the value stated by the producer and the value measured by the use of high-resolution (1m) IKONOS II satellite images. As this variable presents uncertainty in its measurement, it can be represented as interval data. In this way, the efficient frontier is constructed considering the limits of uncertainty (interval limits), that is, the lowest and highest possible values to be assumed by the imprecise variable (a variable with a degree of uncertainty). The proposed theoretical model, called interval DEA model, is presented, as well as the results of the evaluation of the agricultural producers of Holambra using this model.

2. DEA MODEL WITH INTERVAL DATA

The approach proposed here to incorporate interval data to DEA models differs from others in that it makes no assumption in relation to the way in which each input or output varies.

In other words, there is no assumption that the values obey a distribution probability. Whatever the variation, the efficiency frontier is constructed considering only the maximum and minimum possible values to be assumed, with the later use of classic linear programming and traditional DEA models to determine the frontiers.

If there is no certainty about the values assumed for an input or output in a DEA model, equally there will not exist any certainty about the exact location of the efficiency frontier. If the values of the output (input) for some DMUs are greater than assumed, the frontier is moved "higher" ("more to the right"), that is, to a region of greater values for this output (input). If the values are less than assumed, the frontier will be "lower" ("more to the left").

Therefore, in this case, the frontier is not a set in the classic sense of the term, but a fuzzy set and it makes no sense to say that an element does or does not belong to the set; reference should be made to the membership degree of this element to the set. In this way, there will be DMUs with different degrees of membership to the frontier (Soares de Mello et al., 2003).

In this case, the membership degree will be calculated based on the geometric properties of the frontiers generated. So, it is necessary to define some terms. In the case of an output with uncertainty, the frontier located "higher up" is, in fact, the frontier obtained by a classic DEA model, CCR or BCC, which takes into account the maximum value of the uncertain output that each DMU can achieve. As, in terms of production, this is the best situation for all the DMUs, the frontier thus obtained is called the Optimistic Frontier; the frontier obtained by the classic DEA model which considers the smallest value of output for each DMU is the Pessimistic Frontier.

Figure 1 (a) and (b) illustrates these concepts for the model DEA BCC. The fuzzy frontier is the entire region situated between the pessimistic frontier (lower frontier (a) or to the left (b)) and the optimistic frontier (higher frontier (a) or to the right (b)). Note that a DMU is longer represented by a point; the uncertainty in the measurement of the output (a) or input (b) makes the representation of the DMU become a segment of a straight line with limits determined by the pessimistic and optimistic values of this output or input. Note that the segment of straight line, which represents an uncertain input, is horizontal, in contrast to the case of outputs where the DMU is represented by a vertical segment.

In Figure 1, PO_o and PO_p correspond, respectively, to the projected output in the optimistic and in the pessimistic frontier; IP_o and IP_p are the projected input projected in the optimistic and in the pessimistic frontier. c is the length of the DMU, in other words, it is the difference between the optimistic and pessimistic values of the output or input; l is the width of the band, that is, it represents for each DMU, the difference between the value of the uncertain output or input for the optimistic and pessimistic frontiers; p is the part which is inside the band, it is the difference between the optimistic output or input of each DMU and the intersection of this DMU with the pessimistic frontier.

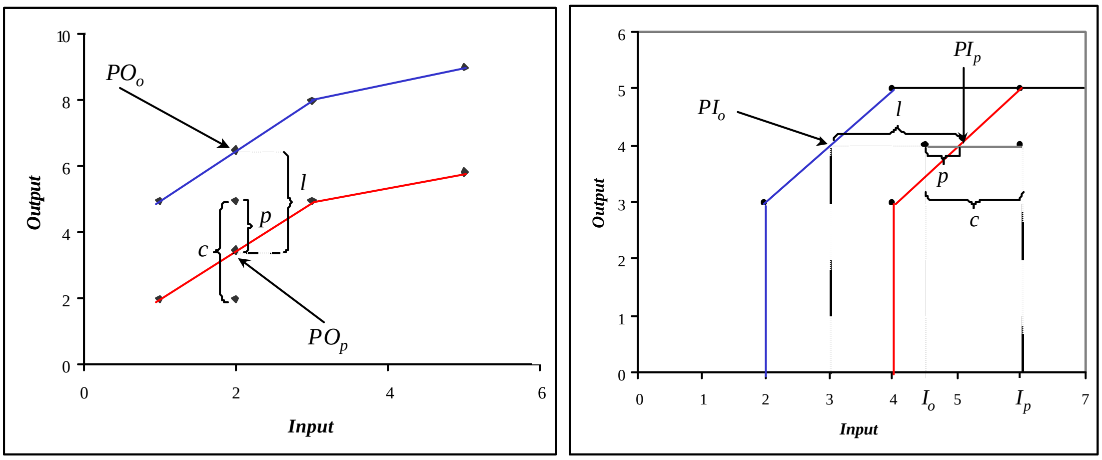
When the fuzzy frontier and the terms c, l and p have been defined, the membership degree of each DMU to this frontier must be defined. In Figure 2 (the case of an imprecise output) it can be observed that the DMUs *A* and *F* are integrally contained in the region that defines the fuzzy frontier. These DMUs must have membership degree 1 to the frontier. The DMUs *B* and *C* however only touch the frontier and, therefore, the membership degree is zero. Between these two extreme cases, the DMUs could have intermediate membership degrees.

The DMU *G* contains all of the width of the fuzzy frontier, but has a part of the width of the band outside the frontier. In other words, in the hypothesis of pessimistic outputs in classic DEA, the DMU would not be efficient. Thus, in spite of not being totally excluded from the frontier, its membership is also not total. In situations similar to that of DMU *G*, membership must obey the relation p/c, unitary when $p=c$.

On the other hand, observation of DMU *E* shows that it is totally contained within the fuzzy frontier, although there is a region of the frontier that does not contain the DMU. Thus, if they are considered optimistic outputs in classic DEA, the DMU is not efficient. For situations similar to this, membership must be p/l, unitary when $p=l$.

It becomes necessary to combine the two cases, in order to guarantee that a DMU only has membership 1 to the frontier if it is efficient both in the pessimistic and in the optimistic hypothesis. The product of the expressions used beforehand, considered partial membership, satisfies this property. In this way, membership to the fuzzy frontier is defined by the equation (I) (Soares de Mello et al., 2003).

$$\wp = \frac{p^2}{lc} \qquad \text{(I)}$$



(a) (b)

Figure 1: Optimistic and pessimistic frontier for interval output (a) or interval input (b).

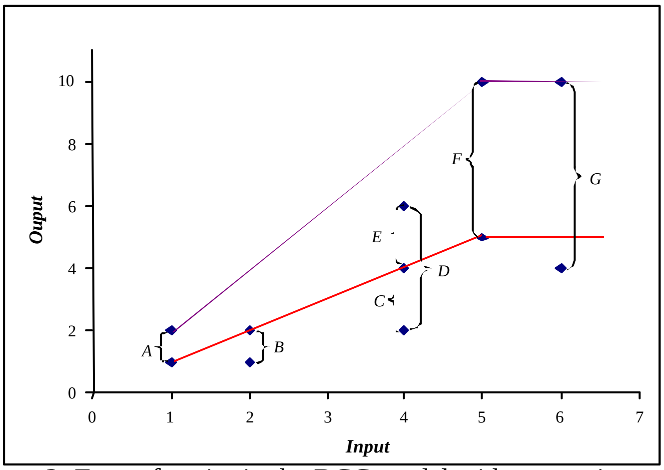


Figure 2: Fuzzy frontier in the BCC model with uncertain output

2.1 Algebraic calculation of membership

For the case of an output with uncertainty, considering the classic definitions of DEA concerning outputs, and that, in this situation, the efficiencies are given by numbers greater than one unit, there are the equations (II) and (III), in which O_o and O_p are the values in the optimistic and pessimistic frontiers of this output; Eff_p and Eff_o are the efficiencies in relation to the pessimistic and optimistic frontiers.

$$Eff_p = \frac{PO_p}{O_p} \qquad \text{(II)}$$

$$Eff_o = \frac{PO_o}{O_o} \tag{III}$$

The width of the band l is the difference between the target of the optimistic frontier and the target of the pessimistic frontier, in other words, $l = PO_o - PO_p = O_oEff_o - O_pEff_p$. The length of the DMU c is the difference between the optimistic output and the pessimistic output, that is, $c = O_o - O_p$. The part of the DMU, which is within the frontier p, is the difference between the optimistic output and the target of the pessimistic output in the pessimistic frontier, provided the difference is positive. This implies that the optimistic output must be inside the band of the fuzzy frontier; if not, p must be equal to 0. In (IV) the equation for p is formulated.

$$\begin{aligned} &p = O_o - O_pEff_p, \quad \text{if } O_o - O_pEff_p \geq 0 \\ &p = 0, \quad \text{otherwise} \end{aligned} \tag{IV}$$

From the previous relationships, it is possible to deduce the expression that represents the membership algebraically. This relationship is presented in (V).

$$\begin{aligned} &\wp = \frac{(O_o - O_pEff_p)^2}{(O_oEff_o - O_pEff_p)(O_o - O_p)}, \quad \text{if } O_o - O_pEff_p \geq 0 \\ &\wp = 0, \quad \text{otherwise} \end{aligned} \tag{V}$$

Similar deductions to those of the case with interval output, permit the definition of the membership score for the case of input with uncertainty, as presented in (VI).

$$\begin{aligned} &\wp = \frac{(I_pEff_p - I_o)^2}{(I_pEff_p - I_oEff_o)(I_p - I_o)}, \quad \text{if } I_pEff_p - I_o \geq 0 \\ &\wp = 0, \quad \text{otherwise} \end{aligned} \tag{VI}$$

3. CASE STUDY: EVALUATION OF THE EFFICIENCY OF AGRICULTURAL PRODUCERS OF HOLAMBRA WITH DEA MODEL FOR INTERVAL DATA

3.1. Agriculture in Holambra

The municipality of Holambra is situated 145 km from the city of São Paulo (the capital of the State of São Paulo, Brazil), in the central eastern region of the State, on latitude south 22º37'55" and longitude west 47º03'36". It has an area of approximately 65 km^2 at an average altitude of 600 m. It has a population of 7,211 according to the demographic census of the Brazilian Institute of Geography and Statistics of the year 2000.

The municipality is characterised by the Dutch immigration to Brazil (at the end of the 40s) and has an economy strongly based on combined agricultural cultivation and stock raising, with a predominance of horticultural activity, citrus fruit growing, decorative plant and flower production, pig farming, poultry farming and dairy farming. It is relatively small in terms of territory in relation to other Brazilian municipalities and almost all of its agricultural production comes from a group of roughly 287 producers, with an average of 20 ha.

Mangabeira (2002) mapped the use of the land in the municipality of Holambra using as a base IKONOS II satellite images (from the year 2000) and field studies (in 2001). Using field study questionnaires, he gathered information on 266 socio-economic variables (which created 204 indicators), from a sample of 74 rural producers.

Having these indicators, the author classified the sample of producers. Six types of agricultural producers were identified in Holambra (Mangabeira, 2002): Citrus Fruit Growers (14 producers); Flower producers (7); Combined Crops and Stock Raising (4); Flower and Citrus Fruit Producers (24); Floriculturists (13); Annual Crop Producers (9).

3.2. Structuring of the problem

For the evaluation of the Holambra agricultural producers, were selected 4 of the 206 indicators deriving from Mangabeira (2002). These variables indicate the classic relationships between production, capital and work. Total Area (represented by the total area of the property) in hectares, Employment (expressed by man-hour-year) and Use of Machinery (calculated as the total number of hours of use of machinery in the year) were selected as inputs; Net Annual Revenue of the property, in Brazilian reais is the output.

The values of the variables, Employment, Use of Machinery and Net Revenue are those supplied by the agricultural producers in the field study questionnaires. The Total Area variable can be obtained from two sources: from the field study questionnaires (in other words, from information supplied by the agricultural producers) and from the satellite images of the municipality. The use of the IKONOS II satellite images combined with the knowledge of the borders of each property permitted the calculation of the total area of each property. This value, however, is different from the value supplied by the agricultural producer. Thus, the Total Area variable is a imprecise variable which can be represented as interval data, one being the limits of the interval to the information given by the agricultural producer and the other, the data calculated based on the IKONOS II satellite image.

The evaluation units (DMUs) are the agricultural properties sampled by Mangabeira (2002). Of the 74 producers, 3 were discounted from this analysis for two reasons: they are milk producers and they present a negative value for the Net Revenue Variable.

The model selected was the DEA BCC model oriented to input, as there are significant differences of scale among the DMUs and the interval variable is input. All the DEA model results necessary for the calculation of the membership score (VI) were obtained with the use of the SIAD software (Angulo-Meza et al., 2003).

3.3. Results

3.3.1. Calculation of the membership score

Both in relation to the optimistic and pessimistic frontiers, 10 DMUs were DEA BCC efficient, being the same in both cases.

When calculating the score $\wp$ the results obtained were 9 DMUs with unitary membership (1 annual crop producer, 1 floriculturist, 1 flower and citrus fruit producer, 2 combined crop producers and stock raisers and 4 flower producers) and 62 had zero membership to the diffuse frontier. There were no DMUs with intermediate membership. This result corroborates the thesis that the flower producers are more efficient than those producers who are involved in other activities (livestock raising, annual crop cultivation etc.).

Due to the large number of ties, it is necessary to discriminate between the units under evaluation more. To do this, the use of the inverted frontier is proposed which, when making the pessimistic evaluation, permits the calculation of a composed evaluation score.

3.3.2. Inverted frontier and composed evaluation score

The membership degree to the frontier is not a measurement of efficiency. In fact, two DMUs that have a zero degree of membership to the frontier can have very different relative positions not detected by the index proposed here. In other words, contrary to the DEA classic models, which supply many ties in the 100% efficient scores, the approach, presented in this article supplies ties for the DMUs with no membership of the frontier.

To distinguish between these DMUs it is necessary to introduce the concept of the inverted frontier (Yamada et al., 1994; Novaes, 2002), which consists of considering the outputs as inputs and the inputs as outputs. An inverted fuzzy frontier can be used to distinguish between the DMUs with a zero membership degree to the original fuzzy frontier.

To obtain a single classification score, the two membership degrees must be encompassed and the variation of the scores made to be between 0 and 1. This index will be called *interval DEA composed score* (*IDCS*) and is given by the equation (VII), in which $\wp_o$ and $\wp_i$ are the membership degree to the original and to the inverted frontier.

$$IDCS = \frac{(\wp_o - \wp_i + 1)}{2} \qquad \text{(VII)}$$

3.3.3. Calculation of the interval DEA composed score

When calculating the composed evaluation score presented in (VII), the results showed 6 DMUs with an aggregate score of 1.00; 1 DMU, with a score of 0.95; 6 DMUs, with a score of zero; the other DMUs a score of 0.50.

Among the groups classified by Mangabeira (2002), it can be seen that the groups Flower and Citrus Fruit Growers, Crop and Livestock Producers and Annual Crop Producers present varied membership scores. Half of the agricultural producers classified as Flower Producers have a score of 1.00 and the rest 0.50. The Floriculturists, meanwhile, have a score of zero or 0.50.

This reinforces the feeling that in relation to the other agricultural producers in Holambra, those who produce flowers are more efficient than the other producers. However, it should be highlighted that the idea that producers who use greenhouses are more efficient than those who cultivate flowers in the open air was not confirmed, as the Floriculturist group, of which 100% possess greenhouses, are less efficient than the Flower Producers (79% with greenhouses).

4. CONCLUSIONS

By not deciding a probability distribution for the uncertainties of the variables, nor a fuzzy function for them, the approach proposed in this article for the incorporation of uncertainties to the classic DEA models brings an advantage in relation to the models currently in existence. In addition to this, it is mathematically simple, as the results are obtained through simple algebraic calculations and classic DEA results.

In relation to the case study, two interesting aspects stand out. The first refers to the integration of remote sensing with operational research techniques. The second relates to the results obtained, which confirm the experience of the specialists in indicating the flower producers as the most efficient. However, it found that the agricultural producers who use greenhouses were incorrectly seen as efficient. One possible explanation for this mistake is that the less rigorous evaluations took into account only the cultivated area, not taking into consideration the fact that these rural properties have large amounts of unused land.

5. REFERENCES

Angulo-Meza, L., Biondi Neto, L., Soares de Mello, J.C.B., Gomes, E.G., Coelho, P.H.G. SIAD – Sistema Integrado de Apoio à Decisão: Uma Implementação Computacional de Modelos de Análise Envoltória de Dados. *Relatórios de Pesquisa em Engenharia de Produção*, v. 3, n. 20. 2003. [http://www.producao.uff.br/rpep/relpesq303/relpesq_303_20.doc]

Mangabeira, J.A. de C. *Tipificação de Produtores Rurais Apoiada em Imagens de Alta Resolução Espacial, Geoprocessamento e Estatística Multivariada: Uma Proposta Metodológica*. Dissertation in Rural Sustainable Development Planning, Campinas State University - UNICAMP, São Paulo, Brazil, 2002.

Novaes, L.F.L. *Envoltória Sob Dupla ótica aplicada na avaliação imobiliária em ambiente do sistema de informação geográfica.* Doctoral Thesis in Production Engineering, Federal University of Rio de Janeiro - UFRJ/COPPE, Rio de Janeiro, Brazil, 2002.

Soares de Mello, J.C.C.B., Gomes, E.G., Angulo-Meza, L., Biondi Neto, L. Índice de Eficiência em Fronteiras DEA Nebulosas. *Proceedings of the XXXV Simpósio Brasileiro de Pesquisa Operacional*, Natal, Brazil, 2003.

Yamada, Y, Matui, T., Sugiyama, M. New analysis of efficiency based on DEA. *Journal of the Operations Research Society of Japan*, v. 37, n. 2, p. 158-167, 1994.