PROBABILISTIC COMPOSITION OF CRITERIA FOR SCHEDULE MONITORING

Annibal Parracho Sant'Anna*
Universidade Federal Fluminense (UFF)
Niterói-RJ
tppaps@vm.uff.br

* *Corresponding author* / autor para quem as correspondências devem ser encaminhadas

Recebido em 02/2009; aceito em 05/2010
Received February 2009; accepted May 2010

Abstract

Time is a key factor in management. Along a project execution, keeping the best completion rates, not too slow and not too fast, is a central objective but such a best rate cannot be suitably anticipated in a precise schedule. It must be determined on the job, on a comparative basis. This paper develops an evaluation system involving the measurement of schedule fitting indicators designed to deal with such conditions. This evaluation system is based on a transformation of the data into probabilities of reaching the frontier of best performances that permits precisely composing measurements on correlated attributes. This feature of the system allows for combining criteria evaluated on elementary and on aggregate levels.

Keywords: probabilistic composition; fuzzy sets; data envelopment analysis; performance management; multicriteria decision analysis; decision support systems.

Resumo

O tempo é um fator-chave na gestão. Em muitos casos, manter a melhor velocidade de execução, nem lenta nem rápida demais, é um objetivo central, mas essa velocidade não pode ser adequadamente antecipada em um calendário preciso. Deve ser determinada na prática, sobre uma base comparativa. Este trabalho desenvolve um sistema de avaliação que envolve a medição de indicadores de ajustamento ao calendário concebidos para lidar com essas condições. Esta avaliação é baseada em um sistema de transformação dos dados em probabilidades de atingir a fronteira de melhores desempenhos que permite compor precisamente valores de atributos correlacionados. Esta funcionalidade do sistema permite combinar critérios avaliados no nível elementar e no agregado.

Palavras-chave: composição probabilística; conjuntos nebulosos; análise envoltória de dados; avaliação de desempenho; análise multicritério; sistemas de apoio à decisão.

1. Introduction

Among the variables that must be controlled in management, those measuring the length of time taken to conclude each task are of fundamental importance. But, in many instances, while keeping the best completion rates, not too slow and not too fast, is a central objective, such a best rate cannot be suitably anticipated in a precise schedule. It must be determined on the job, on a comparative basis.

The variability in the duration of each task becomes, then, a major source of complexity in management. In fact there is always some amount of subjectivity in the evaluation of any kind of attribute. For this reason, in recent years, modeling the sources of uncertainty has become a main subject in scheduling. Examples of the importance of such modeling in different fields of application are in Love *et al.* (2008) and Petrovic & Akoz (2007).

In general, the lack of proper information on the degree of uncertainty in uncertain evaluations considerably impairs the use of such evaluations. The need to take uncertainty into account in the primary evaluations as well as making the uncertainty in the final conclusions explicit has always been an important source of criticism towards combined criteria evaluations. See, for instance, Banker (1993) and Selvanathan & Prasada Rao (1994).

The transformation into probabilities of the options being ranked as the most preferable ones, proposed in Sant'Anna & Sant'Anna (2001), opens a way to overcome such difficulties. This probabilistic approach starts by the evaluation according to each particular criterion to be combined. This initial evaluation may be delivered in the most natural scale. A strategy to generate such scale is ranking, allowing for ties and for empty ranks. The randomness in such measurements is treated by considering the observed values as midpoints of statistical distributions modeled by adding stochastic disturbances to them. Then, the probabilities of attaining the first position are computed. These probabilities can be combined into global measures without the need to assign weights to the criteria.

This article studies the application of probabilistic composition to the simultaneous monitoring of isolated and aggregated tasks. A key feature of such context is the possibility of the individual units evaluated interacting inside groups. The individual performances must be evaluated taking group performances into account.

The presence of unaccounted dependence between the variables combined in global indices is another source of criticism to combined evaluations. Indicators of accomplishment of expected results of simultaneous or successive phases of execution of a task are particularly subject to this kind of dependence. If the same attributes are measured in the analysis made to evaluate the performance of isolated units or production cells and their performance as an aggregate, the same disturbances must be present in the formation of both evaluations. In the probabilistic approach, the correlation between such disturbances may be directly taken into account if the global indices are given in the form of joint probabilities.

An example of application to evaluate performances of drivers of a fleet of urban buses considering a goal of increasing the number of passengers is here presented. If the number of passengers served by the bus that a given driver conducts is the only basis for the evaluation, disregarding the number of passengers served by the other drivers, the drivers may develop the practice of reducing speed to pick up passengers that otherwise would take the next bus, with losses to the total number of passengers served by the line.

The text develops as follows. In the next section, the transformation into probabilities of being the best option is described. Section 3 presents the different points of view that may be adopted in combining such probabilities into global evaluations. Section 4 discusses the relations of the probabilistic approach to Data Envelopment Analysis, and Section 5 the problem of correlation between the partial evaluations. Section 6 presents the problem of taking collective evaluations into account. Section 7 develops the schedule variables treatment and Section 8 considers the application of the probabilistic approach to this problem. Final remarks conclude the paper.

2. Probabilities of Being the Best Option

The key computation in the evaluation of probabilistic preferences is the transformation into probabilities of a particular option being the best within a sample. Such probabilities are a natural measure of the decision maker's preferences. Nevertheless, we frequently start with other forms of measurement. The simplest starting point is ranking the options. To measure preferences based on the level or degree of presence of a given attribute, the relative position of the options may be derived from numerical values of costs or distances, for instance. In other situations there is no such quantifiable attribute and the preferences may be given in terms of common language, such as low, moderate or high preference.

The imprecision in the case of qualitative evaluations is usually taken into account by means of representation through fuzzy intervals (Zadeh, 1965), but it is also present in ordinal and cardinal evaluations and can be represented analogously. To compute the probabilities of being the best option all we need is, besides a ranking (with ties admitted as well as different distances between successively ordered options to assure complete generality), a statistical measure of the uncertainty on each position in that ranking. The uncertainty can be always modeled in the framework of measurement with error. The rank of the option (or any other numerical indication of preference) is understood as a position parameter of a statistical distribution. To model the dispersion, the observed range may provide an estimate for a common range for the distributions of the different individuals. Different assumptions on the form may be made to complete the modeling of these probability distributions.

To make the comparisons easier, the probabilities of being the preferred option may be computed with respect to a sample of fixed size, either randomly generated or withdrawn in fixed percentiles of the set of values attributed to the options under evaluation. For instance, this reference sample may be formed by the nine deciles of this set. This reference sample has the advantage of allowing for comparison of the values for the probabilistic preferences obtained to the value 0.1, which will be given to all options in case they are all indiscernible. If there are less than ten options to be compared, fictitious observations may be interpolated in the sample. Similarly, the observed values may be translated to a Likert scale with nine points and the fixed sample formed by the numbers 1 to 9.

The distribution centered at each of these values may be an asymmetric triangular distribution with constant extreme points. In the case of the Likert scale of nine points, these extreme points will naturally have the values 0 and 10. In the general case, to keep the extreme observations as representative of the first and ninth deciles, these extremes may be determined by stretching the extremes of the set of observed values by 1/8 of the sample range.

A normal distribution, with standard deviation derived from the observed range, may be also assumed, as in Sant'Anna (2005). Or a uniform distribution with a range determined in such a way as to allow for all inversions of ranks considered reasonable, as in Sant'Anna (2002).

The probabilities of being the highest value in the sample can be computed by integrating, with respect to the joint density, the probability of the option under evaluation presenting a higher value than that of each other option in the sample. To compute this probability we ought to divide the range into sub-intervals bounded by the values in the sample.

Let us consider, for instance, the case of triangular distributions centered at the observed values and with fixed extremes. Without loss of generality, these extremes may be assumed to be 0 and 1. Denoting by $x_1, \ldots, x_n$, in increasing order, the evaluations according to a criterion X with respect to which the probabilities of preference maximization are being computed and assuming independence between the disturbances that affect the evaluation of different options, the probability of maximization, for the i-th option, will be obtained by adding the results of the integration, along the sub-intervals mentioned in the previous paragraph, of terms of the form

$$\Pi[1-(1-x)^2/(1-x_p)]\ \Pi(x^2/x_q), \qquad (2.1)$$

where the first product is for $p<j$ and the second for $q>j$, p and q different from i, and with j varying from 0 to n, the number of observations in the sample. This integration will be with respect to the triangular density f_i given by

$$f_i(x) = 2(1-x)/(1-x_i) \quad \text{for } i<j \qquad (2.2)$$

$$\text{and} \quad f_i(x) = 2x/x_i \quad \text{for } i>j. \qquad (2.3)$$

The transformation from ranks to probabilities of being the best or the worst option brings an additional benefit, besides those advantages inherent to taking uncertainty into account: this transformation increases the distance between the most important options. Barzilai *et al.* (1987), Brugha (2000), Lootsma (1998), Tryantaphilou *et al.* (1994), among others, present good reasons to prefer nonlinear scales with this form.

3. Combination of Probabilistic Preferences

A way to derive a unique measure of global preference from the probabilities of being preferred according to each criterion consists of treating these probabilities as conditional on the choice of the respective criterion and computing the total probability of preference by adding the products of these conditional probabilities by the probabilities of choice of each criterion. The difficulty in this approach is to determine the marginal probabilities of choice of each criterion. This is especially difficult if the criteria are correlated. This difficulty can be circumvented if it is possible to rank the criteria and model the joint ranks distribution. In this particular case, the probabilities of choice of each criterion may be computed in the same way as the probabilities of preference according to each criterion.

To deal with more general situations, dependence between the criteria may be directly taken into account if the global preferences are determined in terms of joint probabilities of preference according to the multiple criteria. Different joint probabilities may be employed, depending on the point of view adopted. Different points of view may be characterized in terms of choices between extreme positions on two basic orientation axes. These extreme positions are, on one axis, an optimistic versus a pessimistic point of view and, on the other, a progressive versus a conservative point of view.

In relation to the progressive-conservative axis, the evaluator pays attention to the probabilities of maximizing preference. The progressive evaluator looks after options that are the first in excellence, while the conservative evaluator evaluates them by their ability of not being the last. The term 'conservative' is related to the idea of avoiding losses, while the term 'progressive' is related to the idea of reaching higher standards, i.e. improving.

Regarding the optimistic-pessimistic axis, the optimistic extreme consists of considering the satisfaction of only one criterion as sufficient. All the criteria are taken into account, but the composition employs the connective 'or'. The joint probability computed is that of maximizing (in a progressive composition, or of not minimizing in a conservative one) the preference according to at least one of the multiple criteria. On the opposite end, the pessimistic preference looks for options that satisfy every criterion. The connective is 'and'. The joint probability computed is that of maximizing (or not minimizing) simultaneously the preference according to all the criteria. The terms 'optimistic' and 'pessimistic' are related to the idea of trusting that the most favorable or the less favorable criteria, respectively, will prevail.

By combining the positions in the extremes of these two axes, four different measures are generated. Formally, with M_{ij} denoting as before the probability of the j-th option being the most preferred according to the i-th criterion and with m_{ij} denoting the probability of the j-th option being the least preferred according to the i-th criterion, under the hypothesis of independence, the four global basic measures are given by

$$OC(j) = 1 - \pi m_{ij} \tag{3.1}$$

$$OP(j) = 1 - \pi (1-M_{ij}), \tag{3.2}$$

$$PC(j) = \pi (1-m_{ij}), \tag{3.3}$$

and
$$PP(j) = \pi M_{ij}, \tag{3.4}$$

where π denotes the product operator with m terms obtained by varying i over all criteria.

If the criteria are divided into groups and different points of view are allowed in the computation of the joint probabilities within each group, the number of possibilities increases. A natural division of the criteria into groups consists in criteria for which the optimum is higher and criteria for which optimization means reduction. For instance, criteria based on the measurement of advantages or based on the measurement of disadvantages, criteria related to the production of outputs or criteria related to the use of inputs, criteria related to benefits or criteria related to costs, and so on.

4. Probabilistic Composition and Data Envelopment Analysis (DEA)

The probabilistic approach here applied has in common with DEA the feature of deriving the evaluations from distances to the frontier. However, the computation of the probabilities of being the best option generates more robust classifications because it involves comparison to all the options, not only to those in the frontier.

Among the approaches that may be chosen to combine the probabilistic evaluations, the optimistic and progressive approach is the one closest to DEA. If this approach is used, DEA algorithms may also be employed to combine the partial probabilistic evaluations into a final aggregate score.

Generally, the use of DEA in multiple criteria composition is implemented by first identifying inputs and outputs and then constructing an aggregated index. Examples of the use of such approach include Drake *et al.* (2006), Ramanathan (2006) and Vieira Junior (2008). This corresponds, in the probabilistic composition, to divide the criteria into two blocks, one referring to the frontier of the largest values and the other to that of the smallest values.

But the scope of DEA has broadened considerably over the last decade, with procedures based on all the criteria in the same direction, either as benefit or as cost variables, and aggregated, respectively, by a DEA constant inputs or constant outputs model, as developed by Caporaletti *et al.* (1999) or Lovell & Pastor (1999). Cherchye *et al.* (2004) provide a list of this kind of applications.

This trend may be due to DEA's great advantage of not requiring weights for the criteria. Nevertheless, DEA derives weights that are different for each option under evaluation because they depend on the part of the frontier to which the option is closer. Since the ranking derived from DEA scores results from comparisons to different reference options, it may be disputed, especially if there are different levels of importance or variability among the criteria.

Besides, DEA's optimistic approach of offering, for each option, the choice of the most favorable weights may lead to failure to take certain criteria into account in the evaluation of some options. In the case of simultaneous evaluation on individual and group bases, that may result in the individuals with performance above the average being evaluated by their individual scores while those with performance below the average are evaluated by their group scores. In the DEA framework, this may be avoided by constraining the weights on each individual criterion to be kept below the weight given to the same criteria when applied to the clusters. In the probabilistic approach, a direct treatment to this problem is provided by taking the correlation between the criteria into account.

Another frequent criticism to DEA is related to the lack of statistical evaluations. Various efforts have been made to associate confidence intervals to the efficiency scores and test hypotheses about them. Basic issues on this subject are raised in Banker (1993) and Simar & Wilson (1998). With respect to that, an advantage of the transformation into probabilities of being preferred is that it takes uncertainty in measurements into account from the beginning.

5. Dependence Between Criteria

In Sant'Anna (2009), it was verified that the composition of fuzzy logic (Zadeh, 1978) according to the necessity and possibility concepts, which is equivalent to taking, respectively, the minimum and maximum of the membership probabilities, corresponds to an extreme of the correlation between indicators. The formulae for the probabilistic composition, with the notation presented in Section 3, will then be

$$OC\ (j) = 1 - \max_i m_{ij} \qquad (5.1)$$

$$OP\ (j) = 1 - \max_i (1-M_{ij}), \qquad (5.2)$$

$$PC\ (j) = \min_i (1-m_{ij}), \qquad (5.3)$$

and $$PP\ (j) = \min_i M_{ij}, \qquad (5.4)$$

where $\min_i$ and $\max_i$ denote respectively the maximum and minimum along all the criteria of the group of criteria assumed to be dependent.

Thus, in that extreme of maximal correlation that leads to the composition by the minimum or by the maximum, computation of the joint probability will result in a ranking corresponding to the DEA extreme of permitting each option to be evaluated according to the most favorable criterion. The other extreme corresponds to the assumption considered in Section 3 of independence between the criteria.

The ranks derived from these two extreme assumptions constitute information that may be used complementarily. Moreover, correlation structures in an intermediary position may be explored. For instance, a composition approach may be developed employing a subjective contribution of experts to rank the criteria. This will allow for reducing the number of correlation coefficients that must be known in an iterative algorithm, starting with the two most important criteria and introducing a new criterion at a time. Such small number of successive correlation coefficients s may then be estimated.

Independence between criteria applied to isolated individuals may also be assumed and, after computing the joint probability of preference according to these criteria, the criteria related to collective evaluations may be entered successively in the computation. Then, only the small number of correlations present in this second stage will need to be estimated.

A more complex alternative consists of computing the results arising from a larger range of intermediary values for the correlation coefficients and trying to explain the final probabilities as simple functions of these parameters. Determining cut points where the unit chosen as the best changes will provide more useful information for the decision makers than just ranking according to extreme approaches.

6. Modeling Cooperative Attributes

Disregarding individual efforts and evaluating only on the basis of group achievements may leave out of the performance evaluation important drives for improvement. On the other hand, evaluation systems based on the comparison of individual performances may fail to achieve their main objective, that of enhancing global improvement, by fostering competitive practices where cooperation would be a more profitable asset.

For instance, stimulating the productivity in scientific research by offering grants only to researchers presenting, comparatively, the best results in a list of indicators encourages two kinds of attitudes that will harm the development of more complex research activity. The first is the detachment of individual research from the objectives of the institutions where the individuals are located, which should be the real core of the most productive research projects. The second is developing an opposition of each researcher to the success of their peers which compete for the grants reserved for a same research field.

The evaluation system, even when designed to command the assignment of resources to individuals, must take into account environmental variables that affect collectively groups of individuals or are affected by the joint action of such groups. By not considering the environmental conditions affecting the activities of the community where they are located, the evaluator that intends to judge individual productivity may be only measuring individual results attributable to the context where the work is done but not to personal contributions. Sometimes the absolute results are obtained without any productivity of the individual, but rather by efficiently exploring resources made available by external sources on which distribution neither the evaluator nor the evaluated person has any interference.

The probabilistic composition allows for joining, in the same evaluation system, individual and group indicators in such a way that the evaluation of each individual is affected by the group's performance but individual contributions have a significant impact on their particular evaluation. Such a system will combine variables measuring individual attributes and variables measuring the same attributes in aggregate units of evaluation, and will allow for taking into account the positive correlation that may exist between the stochastic components of these variables.

The key feature of this system is then handling the correlation between criteria applied to clusters of options and criteria applied to individual options. A first orientation to model the correlation in this context will be assuming maximal dependence between cluster indicators and the respective individual indicators. Even if not measuring the same feature, cluster evaluations, being more affected by environmental stochastic factors, must be more correlated among themselves and with the individual evaluations than the individual evaluations among themselves. On the other hand, the approach of assuming independence has the advantage of giving more importance to the numerical distances observed.

7. Schedule Variables

To derive from the control variables of a complex enterprise measures of efficiency of teams or contractors specialized in different parts or stages of the enterprise, some of the variables registered must result from monitoring schedule accomplishment. Other indicators may be raised, directly measuring quantitative outputs, evaluating quality and suitability of products delivered or productivity of manpower. But these variables must be evaluated against a counterpart of measurement of variables of, not only length of the whole enterprise, but also of time employed to reach the results of each particular task.

Besides, it may be important monitoring direct effects of meeting the schedule on the execution of tasks. For instance, managing pacing may be crucial to assure constant availability and full time employment of critical equipment and lack of synchrony may radically change the combined effect of simultaneous materials treatments.

Performance with respect to schedule is affected by interactions between tasks designed to be executed simultaneously or successively. For this reason, motivating for cooperation and valuing effects on global results are sometimes the main goals when designing variables to monitor the ability of meeting the schedule.

These interactions also generate a need to evaluate the performance of each individual or team relatively to the others operating in the same environment. Different building schedules generate, for each particular team or enterprise, different possibilities of exploring proper externalities. Thus, translating to a new context patterns observed in different circumstances, of tasks developed in diverse time and geographical areas, may distort the evaluation, as external factors may entirely differentiate the difficulties to be handled.

Evaluating in terms of joint probabilities of reaching the frontier of best performance allows for fully comparative scores of excellence. In the probabilistic framework, it is easy to compose measurements of volume and quality of outcome with time variables. Besides, the possibilities open by the probabilistic approach, of taking into account correlation and combining aggregate with isolated measurements, are particularly useful in comparative and cooperative environments.

A framework to combine schedule and other variables in a probabilistic way may thus become a key tool in such a context. In the next section, a small example is presented that enlightens the main features of such a framework.

8. Schedule Management

In this section a model to combine criteria that apply to individuals and to groups formed by such individuals is developed. The real life situation explored is that of comparing the performances of drivers in a fleet of urban buses. The goal is to determine the values of bonuses to encourage efforts to collect more passengers.

The drivers are naturally grouped by the shifts of the bus line they work for. Three criteria are employed to assess individual performance: an outcome variable, $P_1(D)$, the number of passengers transported by the bus driven by driver D, and two process variables related to the speed kept in various parts of the route to avoid large spacing between two buses of the same line, which would increase the chance of passengers taking buses of competing companies.

To build the process variables, the times each vehicle passes by previously determined points are recorded. The variables, denoted $T_1(D)$ and $T_2(D)$, are the number of arrivals of driver D at each of these points in a time below a pre-established threshold and the number of times the time interval between the vehicle driven by driver D and the vehicle of the same line following it is lower than another pre-established threshold.

The use of the first of these criteria is aimed at avoiding over-speeding and the second at the opposite. As the delay may be due to traffic circumstances beyond the control of the driver, $T_2(D)$ takes this possibility into account by making a comparison with the time of arrival of the next vehicle. If the next vehicle is too close at the recording point, an absence of external factors forcing driver D to reduce speed is assumed.

In addition to the individual variables, an aggregate variable, $P_2(D)$, measures the number of passengers served by all vehicles in the line during the shift of driver D. To avoid the effect of differences between the shifts, the four variables are measured in terms of weekly variation.

Table 1 shows the values of these variables referring to the ten drivers of the three shifts of a unique bus line. The first two shifts, serving morning and afternoon periods of the day, have three buses, while the third, the evening shift, has four. D_{ij} names the i-th driver of the j-th shift.

Table 1 – Bus Line Data.

	T_1	T_2	P_1	P_2
D_{11}	0	-2	5	3
D_{21}	-3	-1	2	3
D_{31}	0	0	-4	3
D_{12}	0	0	4	9
D_{22}	0	0	3	9
D_{32}	0	0	2	9
D_{13}	-1	-1	4	8
D_{23}	-1	-3	3	8
D_{33}	-2	-4	1	8
D_{43}	0	0	0	8

A first exploratory analysis of this data set highlights the best performance variation of the drivers of shift 2, regarding both the final outcome and the process variables. Examining shift 1, a tendency of the first driver to slow down can be noticed, which may be the cause of a tendency of the second driver to speed up. The third driver stands between the established time bounds, what may explain the worse performance in terms of number of passengers transported. In the third shift, the last driver presents a similar performance. On the other hand, the third driver in this last shift presents a poor performance variation in terms of cooperation in meeting schedules without much gain in terms of individual number of passengers served. Figure 1, representing the values of Table 1 with a set of four columns for each driver evaluation according to each criterion, helps visualizing these differences.

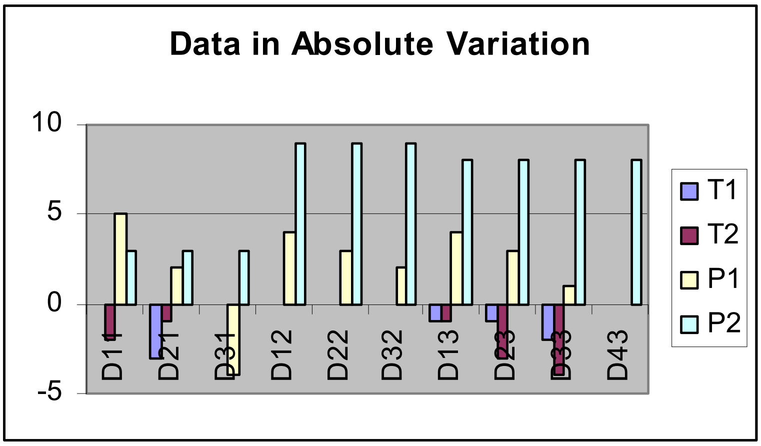

Figure 1 – Representation of the Evaluation of the ten Drivers according to the four Criteria.

Tables 2 and 3, and Figures 2 and 3, show the result of the transformation into probabilities of reaching the upper and lower ranks in each variable, assuming triangular distributions with a range expanded in 1/8 at each end, as described in Section 2.

In Table 3, the high values – above 0.3 – for the probabilities of driver D_{21} minimizing T_1, driver D_{33} minimizing T_2 and driver D_{31} minimizing P_1, as depicted in Figure 3, deserve attention. This is an illustration of the ability of the probabilistic transformation to detach the extreme options.

In Figure 2, the columns representing the values of the probabilities of the most preferred options of Table 2 are not so high. This is due to the ties at the borders. Only the contradiction in the values corresponding to D_{11} outcomes when evaluated in an individual and in a collective basis calls for especial attention.

Table 2 – Probabilities of Being the Best.

	T_1	T_2	P_1	P_2
D_{11}	0.144	0.037	0.215	0.024
D_{21}	0.018	0.064	0.076	0.024
D_{31}	0.144	0.157	0.029	0.024
D_{12}	0.144	0.157	0.146	0.180
D_{22}	0.144	0.157	0.102	0.180
D_{32}	0.144	0.157	0.076	0.180
D_{13}	0.046	0.064	0.146	0.097
D_{23}	0.046	0.026	0.102	0.097
D_{33}	0.026	0.020	0.060	0.097
D_{43}	0.144	0.157	0.050	0.097

Table 3 – Probabilities of Being the Worst.

	T_1	T_2	P_1	P_2
D_{11}	0.057	0.100	0.053	0.241
D_{21}	0.344	0.069	0.078	0.241
D_{31}	0.057	0.053	0.331	0.241
D_{12}	0.057	0.053	0.060	0.036
D_{22}	0.057	0.053	0.068	0.036
D_{32}	0.057	0.053	0.078	0.036
D_{13}	0.083	0.069	0.060	0.042
D_{23}	0.083	0.172	0.068	0.042
D_{33}	0.152	0.327	0.092	0.042
D_{43}	0.057	0.100	0.113	0.042

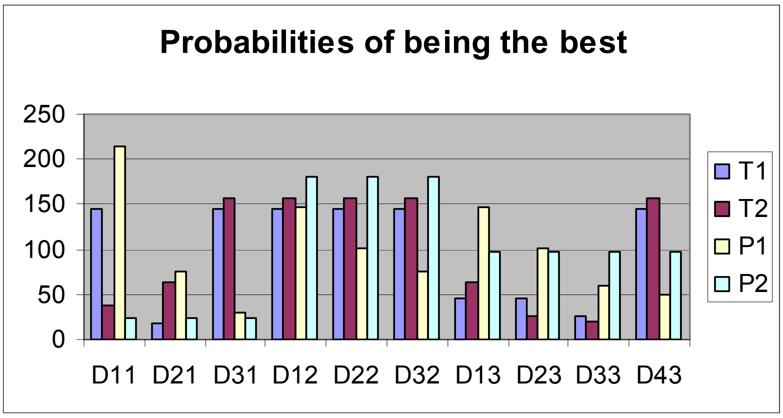

Figure 2 – Probabilities of being the best for each driver according to the 4 criteria.

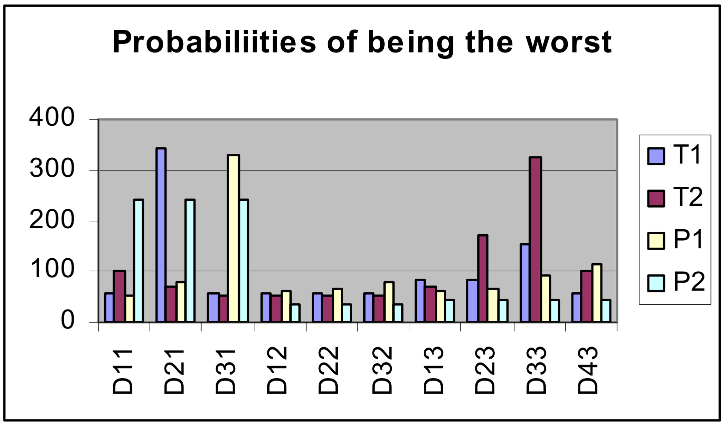

Figure 3 – Probabilities of being the worst for each driver according to the 4 criteria.

Tables 4 and 5 present the probabilistic scores for the compositions under different points of view: OC denotes the optimistic and conservative view and OP the optimistic and progressive one, PC stands for pessimistic and conservative, and PP for pessimistic and progressive.

The values in Table 4 result from modeling as in Section 5 the positive correlation between the aggregate criteria and the related individual criteria. The computation generating, for instance, column OC of Table 4 multiplies the result of application of (3.1) to T_1 and T_2 and of (5.1) to P_1 and P_2. Analogously, for the other columns, the joint probabilities determined by the products of the probabilities of optimizing T_1 and T_2 are multiplied by the minimum between the probabilities of optimizing P_1 and P_2. To generate Table 5, the computations assuming independence, employ, following (3.1) to (3.4), the products of the probabilities of optimizing each of the four variables. To preserve the scale and facilitate comparison, the final scores so obtained are derived from the joint probabilities by taking a cubic root in Table 4 and a fourth root in Table 5.

Table 4 – Scores under Dependence between Outcome Variables.

	OC	OP	PC	PP
D_{11}	0.949	0.070	0.864	0.050
D_{21}	0.917	0.036	0.774	0.030
D_{31}	0.959	0.110	0.842	0.081
D_{12}	0.944	0.149	0.944	0.149
D_{22}	0.941	0.135	0.941	0.132
D_{32}	0.939	0.126	0.938	0.120
D_{13}	0.930	0.070	0.930	0.066
D_{23}	0.901	0.057	0.891	0.049
D_{33}	0.834	0.036	0.803	0.032
D_{43}	0.918	0.118	0.910	0.104

Table 5 – Scores under Independence.

	OC	OP	PC	PP
D_{11}	0.908	0.108	0.884	0.072
D_{21}	0.855	0.046	0.809	0.038
D_{31}	0.876	0.091	0.821	0.063
D_{12}	0.950	0.157	0.949	0.156
D_{22}	0.948	0.146	0.947	0.143
D_{32}	0.946	0.140	0.944	0.133
D_{13}	0.938	0.089	0.936	0.081
D_{23}	0.920	0.068	0.907	0.059
D_{33}	0.882	0.051	0.839	0.042
D_{43}	0.928	0.113	0.922	0.102

For the present problem of motivating to increase outcome, a progressive point of view may be the most appealing. Besides, a pessimistic point of view may be appropriate to call attention to both outcome and process variables. Thus, the last column of the global evaluations is the one to be considered as the basis for the decision to be taken.

However, strong correlations can be noticed between all the vectors of final scores, which present changes only in intermediary ranks. This agreement can be found even between the scores in Tables 4 and 5, demonstrating the robustness of the composition assumptions against the effects of dependence. In fact, changing the assumption from dependence to independence would result in only one significant change in the scores associated with the progressive and pessimistic point of view, with D_{31} falling to a position below those of D_{11} and D_{13}.

These inversions do not follow directly from the changes in the dependence assumptions. They are more related to the greater importance given to the outcome criteria if independence is assumed. To avoid giving higher weight to a group with a larger number of independent criteria, the probabilistic composition may be applied first within the groups, and the product of the probabilities of preference of criteria in the same group replaced by their geometric mean. Slightly different results are obtained if this approach based on separation of the criteria into groups is taken.

If the two process criteria are thought as forming a group independent of the group of outcome criteria and a score of preference according to the latter is composed with the preferences according to the former two criteria, the final scores in Table 6 are obtained. In Table 6, assuming dependence, the two outcome probabilities are composed by the minimum, while the process score is the square root of the product of the probabilities of optimizing T_1 and T_2, which are supposed to be independent random variables.

Comparing the scores in Table 6 to those in the previous tables, again few changes can be noticed. The most important is the better performance of D_{11} under the optimistic and progressive combination in Table 6. This is a consequence of the optimistic stimulus to disregard criteria with respect to which the performance of the agent is worse. With a good performance with respect to P_1, D_{11} is favored by this combination approach. As for D_{31} under the pessimistic and progressive point of view, the score with maximal correlation between the outcome criteria falls from 0.081 in Table 4 to 0.060 in Table 6.

Table 6 – Scores for Equal Group Weights and Dependence.

	OC	OP	PC	PP
D_{11}	0.969	0.156	0.836	0.042
D_{21}	0.946	0.059	0.770	0.029
D_{31}	0.934	0.092	0.795	0.060
D_{12}	0.989	0.165	0.943	0.148
D_{22}	0.989	0.165	0.939	0.124
D_{32}	0.988	0.165	0.934	0.107
D_{13}	0.986	0.102	0.932	0.073
D_{23}	0.982	0.070	0.901	0.058
D_{33}	0.970	0.061	0.828	0.037
D_{43}	0.981	0.124	0.904	0.086

Figure 4 provides a graphic view of the effect of different assumptions on dependence. For each driver, the first column shows the score for dependence of Table 4 and the last column the score for dependence with equal weight to process and outcome variables of Table 6, compared to the score obtained under independence (Table 5) in the middle column. To make fair the visual comparison, in the construction of this figure each of the three vectors of scores being compared is standardized to maximum equal to 1 by dividing the scores by their maxima. It is clear the agreement between the results, with rare inversions of drivers ranks.

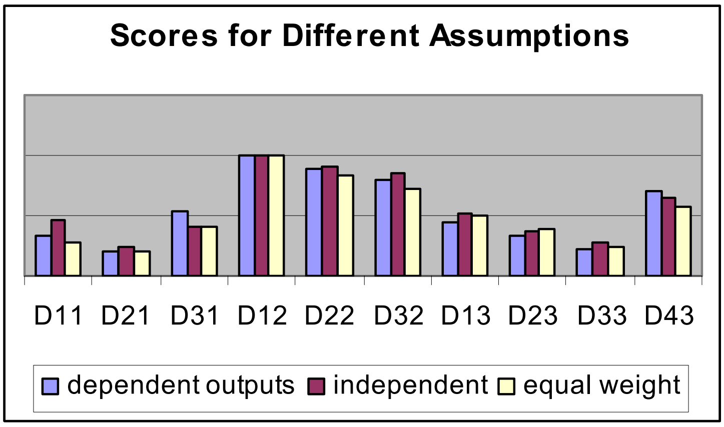

Figure 4 – Comparison of Results of Different Assumptions on Dependence.

To enlighten the discussion developed in Section 4 of the differences between the probabilistic composition and DEA, the software SIAD (Meza *et al.*, 2005) was employed to determine efficiency scores for the 10 drivers of this example. Since the goal intended in the present case is raising quality and not productivity in the use of any inputs, the process variables T_1 and T_2 should be maximized in the frontier of excellence as well as the outcome variables. This may be dealt within DEA by applying an invariant inputs model. The result of the application of a constant returns of scale DEA model with invariant input to the four vectors of outputs given by the probabilities of each of the drivers being the best presented in Table 2 results in six of the ten drivers being efficient. The relatively inefficient drivers are D_{21}, D_{13}, D_{23} and D_{33} with scores of efficiency of, respectively, 0.477, 0.837, 0.542 and 0.539.

The large number of efficient drivers was to be expected, since, for instance, all drivers of a shift would necessarily be efficient in DEA because they will present the maximum value in the fourth variable. This is the main reason why DEA is less suitable to deal with the problem here studied of some variables being measured in a group basis. Besides, combining that with the freedom to choose the most favorable weights leads to the variable T_1 being left irrelevant.

A deeper analysis of DEA results shows also that the efficient drivers by DEA are, as should be expected, those with the best scores in the second column of Table 5, that means, the best according to the optimistic and progressive approach of probabilistic composition if independence is assumed.

Leaving out the problem for DEA brought by the aggregation of evaluated units, comparing the results of the application of DEA with those of the probabilistic composition shows that, in general, while DEA produces scores of simpler interpretation, the probabilistic composition, on the other hand, produces more informative results. Thus DEA may be more interesting for an immediate application and the probabilistic composition more interesting in a system to foster improvement in the long run.

Figure 5 provides a view of the variation along the different composition approaches. In the construction of this figure, as in the generation of Figure 4, the probabilities according to each approach are, divided by their maxima. Besides leaving clear the larger variability of the results along the approaches than along the dependence assumptions, this figure stresses the larger variability of the progressive pessimistic approach.

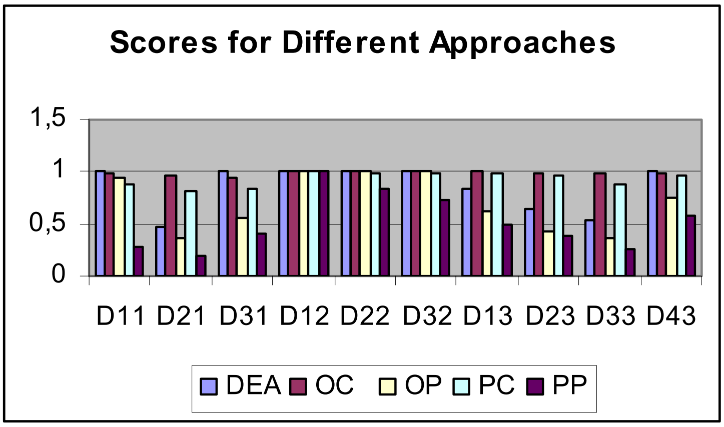

Figure 5 – Comparison of Results for Different Composition Approaches.

It is important to notice, finally, that the constancy of the results is limited not only to the combination point of view chosen but also to the conceptual approach taken. The rankings along the columns of Tables 4 to 6 did not vary significantly when new criteria to globally evaluate the performance of each shift with respect to the process variables already in the analysis were added. But if, on the contrary, the change outreaches the scope of this analysis to take into account other variables like contribution to vehicle maintenance or careful driving, for instance, changes ought to be expected.

9. Final Remarks

The modeling approach developed above shows the suitability of the probabilistic composition of preferences to combine criteria applied on different levels of aggregation. The application developed can be extended to a large number of variables and levels without any conceptual change.

The case studied employs a basic framework for the exploration of dependence relations between criteria. Only extreme dependence relations were assumed. Efforts should be taken to obtain a quantitative basis of information on possible intermediary correlation structures.

The application to other instances of management will bring new opportunities of development. A feature of the evaluation system here developed is its full independence of the availability of precise numerical measurements to start with. This makes it useful in other areas of application where qualitative criteria unrelated to simple quantitative attributes must be considered.

Acknowledgements

The author is grateful to the referees of Pesquisa Operacional for their comments, which resulted in great improvement of the text. This work is supported by a CNPq grant.

References

(1) Banker, R.D. (1993). Maximum Likelihood, Consistency and Data Envelopment Analysis: a Statistical Foundation. *Management Science*, **39**, 1265-1273.

(2) Barzilai, J.; Cook, W. & Golany, B. (1987). Consistent Weights for Judgment Matrices of the Relative Importance for Alternatives. *Operations Research Letters*, **6**, 131-134.

(3) Brugha, C.M. (2000). Relative Measurement and the Power Function. *European Journal of Operational Research*, **121**, 627-640.

(4) Caporaletti, L.E.; Dula, J.H. & Womer, N.K. (1999). Performance Evaluation based on Multiple Attributes with nonparametric Frontiers. *Omega*, **27**, 637-645.

(5) Cherchye, L.; Moesen, W. & van Puyenbroeck, T. (2004). Legitimately Diverse, yet Comparable: on synthesizing Social Inclusion Performance in the EU. *Journal of Common Market Studies*, **42**, 919-955.

(6) Drake, L.; Hall, M.J.B. & Simper, R. (2006). The Impact of Macroeconomic and Regulatory Factors on Bank Efficiency: A non-parametric Analysis of Hong Kong's Banking System. *Journal of Banking and Finance*, **30**, 1443-1466.

(7) Lootsma, F.A. (1988). Numerical Scaling of Human Judgment in pairwise-comparison methods for fuzzy multicriteria decision analysis. *NATO ASI F, Computer and System Sciences*, **48**, 57-88.

(8) Love, P.E.D.; Davis, P.R.; Edwards, D.J. & Baccarini, D. (2008). Uncertainty avoidance: public sector clients and procurement selection. *International Journal of Public Sector Management*, **21**, 753-777.

(9) Lovell, C.A.K. & Pastor, J.T. (1999). Radial DEA Models without Inputs of without Outputs. *European Journal of Operational Research*, **118**, 46-61.

(10) Meza, L.A.; Biondi Neto, L.; Soares de Mello, J.C.C.B. & Gomes, E.G. (2005). ISYDS – integrated system for decision support (SIAD – sistema integrado de apoio a decisão): a Software Package for Data Envelopment Analysis Model. *Pesquisa Operacional*, **25**, 493-503.

(11) Petrovic, D. & Akoz, O. (2007). A Fuzzy Goal Programming Approach to Integrated Loading and Scheduling of a Batch Processing Machine. *The Journal of Operational Research Society*, **59**, 1211-1219.

(12) Sant'Anna, A.P. & Sant'Anna, L.A.F.P. (2001). Randomization as a Stage in Criteria Combining. **In**: *Production and Distribution Challenges for the 21st Century* [edited by J.L.D. Ribeiro, F.S. Fogliatto and L.B.M. Guimarães], ABEPRO, Salvador, 248-256.

(13) Sant'Anna, A.P. (2002). Data Envelopment Analysis of Randomized Ranks. *Pesquisa Operacional*, **22**, 203-215.

(14) Sant'Anna, A.P. (2005). Composição Probabilística de Critérios na Avaliação de Cursos. *Revista Brasileira de Pós-Graduação*, **2**, 40-54.

(15) Sant'Anna, A.P. (2009). Probabilistic Majority Rules: a New Approach to the Composition of Social Preferences. *International Journal of Industrial and Systems Engineering*, **4**, 19-31.

(16) Selvanathan, E.A. & Prasada Rao, D.S. (1994). *Index Numbers: A Stochastic Approach.* University of Michigan Press, Ann Arbor.

(17) Simar, L. & Wilson, P.W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science*, **44**, 49-61.

(18) Tryantaphilou, E.; Lootsma, F.A.; Pardalos, P.M. & Mann, S.H. (1994). On the Evaluation and Application of Different Scales for Quantifying Pairwise Comparisons in Fuzzy Sets. *Journal of Multicriteria Decision Analysis*, **3**, 133-155.

(19) Vieira Jr., H. (2008). Multicriteria Approach to Data Envelopment Analysis. *Pesquisa Operacional*, **28**, 231-242.

(20) Zadeh, L.A. (1965). Fuzzy sets. *Information and Control*, **8**, 338-353.

(21) Zadeh, L.A. (1978). Fuzzy Sets as the Basis for a Theory of Possibility. *Fuzzy Sets and Systems*, **1**, 3-28.