NEW EVIDENCE ON THE FORMATION OF PREFERENCES OF BETTERS IN HORSE RACES

Annibal Parracho Sant'Anna (UFF) aparracho@uol.com.br
Luiz Flávio Autran Monteiro Gomes (IBMEC) luiz_autran_gomes@terra.com.br

Resumo

Este artigo apresenta evidência adicional sobre a contribuição das probabilidades de ser a melhor escolha de acordo com critérios particulares para a formação de preferências dos apostadores em páreos selecionados de corridas de cavalos. São consideradas regras de combinação baseadas na projeção sobre o vetor de avaliações da alternativa em maior evidência, na distância à fronteira de excelência e em abordagens probabilísticas. Dois critérios são combinados: a opinião de especialistas e a escolha dos jóqueis. A análise é baseada em dados das provas de grandes prêmios disputadas durante a semana do Grande Prêmio Brasil de 2002.

Palavras-chave: preferências – probabilidades de escolha – análise de decisão multicritério

Abstract

This paper presents further evidence concerning the contribution of the probabilities of being the best choice according to particular criteria to the formation of preferences of the betters in selected horse races. Combination rules based on projection on the vector of evaluations of the option under strongest evidence, on distance to the excellence frontier and on probabilistic approaches are considered. Two criteria are combined: experts' opinion and jockeys' choice. The analysis is based on data from the big prize races run during the week of the Grande Prêmio Brasil 2002.

Key words: preferences - probability of choice - multicriteria decision analysis

1. INTRODUCTION

This paper deals with the composition of preferences derived from ranking according to multiple criteria. The basic model explains the final probabilities of choice in terms of the preferences elicited when judgement is based on each criterion in isolation.

Evidence was collected in Sant'Anna (2002a) that the application of the replacement of ranks by probabilities of being the best choice helps to explain the formation of betters' preferences in horse races. Two criteria were considered there: preferences derived from the past performance of the competitors, provided for special races by the Official Track Program, and preferences exhibited by the jockeys as they chose their mounts.

In this article, attention is centred on the process of combining the criteria to derive the final preferences ordering. The following section discusses the derivation of the probabilities of being the best choice. In Section 3, criteria combination is discussed. Section 4 studies the formation of preferences in horse races. Finally, in Section 5, the proposed approaches are applied to the data of the big prize races run, at the Brazilian Jockey Club in Rio de Janeiro, during the week of the Grande Prêmio Brasil – 2002.

2. PROBABILITIES OF CHOICE

The preference must be based on objective knowledge. However, its construction passes through psychological channels. In many situations, the distribution of preferences observed deviates from what would result from the rational application of the information and theoretical models available to the decision-makers. Here we propose a way to model one of these psychological channels by following the approach of Gomes and Lima (1992) of modifying distances between options.

We focus on one factor: the need to simplify things. The tendency of the human mind to follow simple decision rules is the fundamental premise of the theory named Nomology by Brugha (1998). This leads to an attitude which disregards numerical values when dealing with small probabilities and huge gains, fitting the management of loss aversion and risk propensity on different points of the value scale of Prospect Theory (Kahnemann and Tversky, 1979). This behaviour is here condensed in the transformation of the initial evaluations according to each criterion into probabilities of being chosen according to such a criterion.

We take as initial information a full set of coherent preference classifications according to well-defined criteria. Setting the evaluations in terms of ranks makes the process easier to follow. The necessary coherent preference indications from which the ranks are derived may be obtained by attributing to each option a value in a previously defined scale where each point corresponds to a position that is easy to determine according to the criterion that is being applied. We may also start by comparing pairs of alternatives, according to such criterion, like Saaty (1980) or Lootsma (1993).

One way to simplify the choice of the prospective winners in horse races is to stretch the distances between the best options and the others in such a way as to practically disregard options of lower chance. A well-known way to transform the variables in the direction of expanding distances between the preferred options is to change from an arithmetic to a geometric scale (Brugha, 2000). This may be done by treating the ranks as exponents on a given basis. In this way, ranks are transformed into probabilities with a geometric distribution.

Instead of transforming the variables, a probability distribution is directly introduced here, by means of a systematic mechanism fully developed in Sant'Anna (2002b). It consists of introducing a random component into the model for the preference. With the addition of this random component, the preference, initially postulated in a deterministic fashion, comes to be seen as an estimate of the mean of a probability distribution. By treating the indications of preference as observations of random variables the probability of each option taking the position of highest preference can be calculated.

To maximise the possibility of options classified close together to change their ranks, we impose the uniform distribution to the random components of the preference measures. In the uniform family, the distribution around the expected value is perfectly identified by the information on a dispersion parameter. If we wish to allow any two options to invert their relative positions, the range of the distribution must be larger or equal to the difference between the initial highest and lowest preference values. If the preferences are given in terms of ranks, this difference equals the number of available options less 1.

Formally, the transformation applied to the ranks will then consist of replacing the rank R_{ij} of the j-th option according to the i-th criterion by the probability that, according to this criterion, the option would be placed in the position of highest preference, under the assumption that, for all i and j, the preference by the j-th option according to the i-th criterion is a random variable uniformly distributed around the respective register R_{ij}. In addition, these uniform distributions are assumed to be independent, with all those relative to the same criterion endowed with the same range parameter. This parameter is determined, for the i-th criterion, by the maximum of the differences $R_{ik} – R_{il}$, for k and l varying along all the options evaluated.

3. APPROACHES TO THE COMBINATION OF CRITERIA

There are different systematic ways of combining the probabilities of choice according to the different criteria. The tendency to prefer simpler representations also affects the composition of the criteria through the prominence effect, which leads to disregarding evaluations according to criteria that affect the most probably chosen options less.

At this point a way to bring the prominence effect into the composition procedure is established. It consists of first choosing the main criterion from among the multiple evaluation criteria employed and singling out the vector of probabilities of choice of the option with the highest probability of choice according to this main criterion. Weights are then assigned to the criteria, proportional to the probability of choice of the preferred option according to the main criterion. This is equivalent to taking as the final measure of preference for an option, the norm of the projection of the vector of probabilities of choice of this option, according to the multiple criteria, on the direction determined by that option preferred according to the main criterion.

This procedure is asymmetric in the sense of involving preference among the criteria. Symmetric probabilistic and deterministic combination rules may be set. Some of these are compared here. Two different probabilistic rules are given by the probability of being the best choice according to all the criteria or according to at least one of them. Asking for the best choice according to all the criteria is more capable of stretching the distance between the preferred option and the others when there is an option preferred according to most criteria. A third probabilistic composition rule operating in the opposite direction, reducing distances, would be to measure the preference in terms of the probability of not being classified as the worst option according to all criteria.

Besides these probabilistic combination rules, a symmetric deterministic rule is also considered. It is given by the minimum distance to the frontier determined by the best choices. We compute it through the DEA (Charnes, Cooper and Rhodes, 1978) efficiency score, considering all options as production units, all of them provided with the same amount of a given input and presenting as output the probabilities of choice according to the different criteria. The DEA principle of offering each unit the more favourable weights makes this option similar to the probabilistic rule based on determining the preference according to the probability of being chosen by at least one criterion.

4. THE CASE OF HORSE RACES

The modelling strategy developed above is applied here to explain the formation of preferences of betters in horse races. In this case, we know the final preferences distribution precisely, given by the amount of money bet on each animal. The moment when the race starts captures a probability distribution of bets that faithfully mirrors the group's opinion on the horses' chances of winning.

This probability distribution does not need to represent the reality, in the sense that only random disturbances, intervening during the race to modify it, would determine the winner. The process of formation of the betters' preferences may not take into due account factors that systematically affect the results of the races. Such factors may be out of the range of knowledge of the betters who may also form their preferences on the basis of erroneous theories. It is also possible that emotional factors deviate in the same direction from the objective of taking advantage of the possible distortions in the observed cost/benefit ratio of each possible bet.

Among these factors, are, on one side, the convenience of substantiating the choice on a simple comparison and, on another side, the unreliability inherent to the result of any rational procedure of choice, given the aspects that people are always forced to leave outside any analytical model. The conjugation of these two principles would cause a concentration of the bets in the options of higher probability and a calculation of the chance of the other options with regard to those. If we rank from the least preferred to the most preferred, giving rank one to the worst option, and then replace such ranks by the probabilities of each option presenting the highest rank, we come to obey these two principles.

The replacement of ranks by the probabilities of being the best choice is applied here to two orderings: the preferences derived from the past performance of the competitors, supplied by an expert in the Track Program, and the preferences exhibited by the best jockeys as they choose their mounts. These are important criteria for the betters, but to produce more realistic models, multiple, rather than two, criteria should be combined. Since all the modelling alternatives studied extend easily to more than two criteria, we use the simplest model in this example.

The starting point is given by the ranks of the competitors according to each criterion. These are considered as observed values of random variables and estimates of their expected values. From the joint distribution of these random variables, built by applying hypotheses about the form of the distribution and hypotheses of independence and identical dispersion, the probability of each option being the preferred choice is derived.

If we were able to rank the options globally, this procedure might be applied directly to the global ranks to derive final preferences. However, this is not usually the case and, after obtaining the probability of being the best choice according to each criterion, we still have to combine these partial evaluations. The principle of prominence leads us to combine them through the comparison with the most evident options.

In our case, the first criterion, preference provided by the Track Program is a more reliable criterion than the preferences signalled by the jockeys. These are bound by long-term relationships with trainers and owners. In such a situation, the reference option will be that preferred according to the Track Program expert. Then we reduce the pair of measures of preference for each animal according to the different criteria to a final one-dimensional measure by projecting the vector of preferences for each horse on the direction given by the vector of preferences by the animal preferred according to the first criterion.

5. APPLICATION

The results of the transformation on probabilities of being the best choice are combined here in different ways to explain the preferences of the betters in a set of big prize races. The objective of this investigation is to compare probabilistic combinations with the combination through projection and through distance to the excellence frontier. The basis for this comparison is the correlation to the observed final bets. The correlation with the preferences vector derived from each criterion is compared to the correlation with the additive combination with the best coefficients, determined through the least squares optimisation principle.

The application of the projection on the direction found the most important results in a correlation to the vector of bets very close to that obtained by the best linear adjustment. This provides statistical support for the conjecture that the mechanism of aggregation involves projection on the direction of the option preferred by the main criterion and for the use of the norm of the vector of projected preferences.

The table below presents the correlation, corresponding to each race examined, of the final probability of being the best choice, as constructed by the betters, with each preference vector. In the probabilistic combinations, the effect of the correlation between the criteria was neglected. This is a common practice justified, in the present case, because assuming the correlation for all the options to be identical would force to accept null correlation estimates.

The total referred to in the last line of this table is for the nine races. We can see, in the total, a correlation to observed bets of 83% for the vectors of preferences determined by the norm of the projection of the vector of probabilities of being the best choice according to the two criteria on the direction determined by the probabilities of the option preferred according to the dominant criterion. This correlation is identical to that provided by the best empirically determined linear combination.

Race	LEAST SQUARES	NOT WORST	BEST IN SOME	BEST IN BOTH	DEA	PROJECTION
Brasil	94%	39%	86%	83%	81%	94%
Breno	81%	30%	76%	94%	59%	81%
Cidade	89%	41%	85%	89%	80%	89%
Delegações	87%	42%	80%	94%	63%	85%
Mossoró	92%	49%	88%	97%	80%	92%
Presidente	94%	31%	83%	81%	79%	93%
Seabra	84%	43%	76%	98%	60%	83%
Sukow	76%	25%	61%	58%	55%	72%
Tirolesa	68%	1%	50%	21%	47%	67%
Total	83%	31%	77%	79%	64%	83%

TABLE 1. 2002 CORRELATIONS TO OBSERVED PREFERENCES

6. FINAL COMMENTS

In this work we have sketched some alternatives for handling different factors affecting the formation of preferences. We have demonstrated that models based on the calculus of the probability of being the best choice are able to identify behaviour suggested by the theory. The projection on the direction in main evidence is also shown to improve the explanation of the betters' final behaviour.

The tips presented in the racing form by the expert do not mention all the animals. Those not mentioned were left tied in the position of worst choice. This performs part of the job of the transformation into probabilities of being the best choice. Even so, data analysis offers support to the idea that the prominence principle would act through the projection on the preferred option.

The results obtained are clear in the context of formation of preferences of betters in horse races. The idea of combining probabilities of choice according to particular criteria to explain global probabilities of choice is suitable to any context. To extend the full strategy to other contexts merely requires identifying a dominant criterion and verifying the assumption that the motivation to take the best option as reference is, in fact, present.

REFERENCES

Brugha, C. M. (1998) - The Structure of Qualitative Decision Making, *European Journal of Operational Research*, **104**, 46-62.

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

Charnes, A., Cooper, W. W. and Rhodes, E. (1978) - Measuring the Efficiency of Decision Making Units, *European Journal of Operational Research*, 2, 429-444.

Gomes, L. F. A. M. and Lima, M. M. P. M. (1992) - From Modelling Individual Preferences to Multicriteria ranking of Discrete Alternatives: a Look at Prospect Theory and the Additive Difference Model, *Foundations of Computing and Decision Sciences*, 17, 172-184

Kahnemann, D and Tversky, A. (1979) - Prospect Theory: an Analysis of Decision under Risk, *Econometrica*, **47**, 263-291.

Lootsma, F. A. (1993) - Scale Sensitivity in the Multiplicative AHP and SMART. *Journal of Multicriteria Decision Analysis*, **2**, 87-110.

Saaty, T. L. (1980) - The Analytic Hierarchy Process, Planning, Priority Setting and Resource Allocation. Mc. Graw Hill, New York.

Sant'Anna, A. P. (2002a) - Randomisation and Composition of Preference Measurements, *Pesquisa Operacional*, **22**, 87-103.

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