Game theory concepts and changes in the Brazilian agriculture

Fernando L. Garagorry

Embrapa, Secretaria de Gestão e Desenvolvimento Institucional
Ed. Sede, Caixa Postal 08815, 70770-901 Brasília, DF, Brasil
fernando.garagorry@embrapa.br

Abstract

The evolution of the Brazilian agriculture, over the last four decades, has shown substantial changes in the geographical distribution of individual products, land use, production value and many other variables. Besides, even within a fixed geographical area, many changes have taken place. The study of these variations is based on large amounts of available data and the use of non-traditional analytical techniques, in order to obtain new information. One of these techniques, based on the power indices of game theory, has been applied to datasets on individual products, land use and production value.

1 Introduction

The Brazilian agriculture has shown, over recent decades, substantial changes in the geographical distribution of products, land use, manpower, machinery, production values and many other variables. The study of this kind of motion over the national territory is being conducted under the general name of "agrodynamics" [2]. Following the official Territorial Division of Brazil, the studies have considered the levels of the whole country (1), regions (5), states (27), mesoregions (137) and microregions (558), for a total of 728 geographical entities. Nevertheless, many data are available down to the level of municipalities (their number has changed over the years, and presently is 5570). Besides the changes in geographical terms, other variations have taken place within a fixed territorial entity. In any case, many types of changes can be identified and assessed with the same techniques.

The following situations have been considered: a) distribution of the "volume" (i.e., stock headcount for animals or produced quantity for all the other items) of individual products in the five regions, with annual data, many going back to 1975; b) distribution of agricultural area into six types of land use; c) distribution of monetary value of agricultural production into eigth classes. In the last two cases, the data come from six agricultural censuses. In all these situations, the main analytical techniques are based on a simple step: in any year, the original values in the different classes are divided by the total of the respective additive variable and a relative distribution is obtained. In the paper, the numbers are multiplied by 100, so that the situations are presented as sets of percent distributions.

Since a distribution can also be seen as a voting game, it is possible to determine some indices of power for that game. Here, only the Banzhaf-Coleman and the Shapley indices will be considered. There are some important things to be mentioned about these indices: 1) they have been presented as "solution concepts" for cooperative games; 2) the same index of power can correspond to different games; 3) usually, the indices of Banzhaf-Coleman and of Shapley are not very different; 4) the Banzhaf-Coleman index is easier to evaluate than that of Shapley. The main idea is to use such indices to discriminate between seemingly close distributions or to group seemingly quite different ones. That is, the use of power indices appears as an auxiliary data mining technique.

2 Indices of power

An introduction to the theory of n-person cooperative games can be found in Owen [3, Chapters 8-11]. We consider an n-person game, with $n > 2$, which has the form of a simple majority voting game. N will denote the set of players. Throughout this paper, the original values of the players will be presented as an ordered set of n non-negative numbers, which add up to 100. The value of a coalition is the sum of the values of its members. A winning coalition is anyone whose value is greater than 50. A *swing* for player i is defined as a set $S \subseteq N$ such that $i \in S$, S wins, and $S - \{i\}$ loses.

Letting b_i be the number of swings for player i, the following number can be calculated:

$$B_i = b_i / \sum_{j=1}^{n} b_j \ . \tag{1}$$

This is the normalized Banzhaf-Coleman index for player i. More generally, the Banzhaf-Coleman index of power, for the original game, is the list $B = (B_1, B_2, \cdots, B_n)$. By construction, this is a list of non-negative numbers which add up to *1*; in this presentation, these numbers will be multiplied by *100*. For simplicity, the list will be referred to as the Banzhaf index of power. A detailed discussion of the mathematical properties of this index was presented in Dubey and Shapley [1].

In a similar way, at least for the type of games being considered here, the Shapley value for player *i* can be defined as follows:

$$S_i = \sum_T \frac{(t-1)!(n-t)!}{n!} \ , \tag{2}$$

where the summation is taken over all swings T for player i, and t is the number of players in T. The Shapley value for the original game is the list $S = (S_1, S_2, \cdots, S_n)$. It is well known that this is a list of non-negative numbers which add up to *1*; in this presentation, as before, these numbers will be multiplied by *100*.

Therefore, each of the above lists, that is the Banzhaf index or the Shapley value, can be taken as defining a new simple majority game, which was deduced from the original game. The important thing is that they capture some essential facts relating to the winning coalitions of the original game. For some games both lists coincide, but this is not always the case.

3 L1 distance between two distributions

As presented above, a game can also be seen as a percent distribution; that is, a set of non-negative numbers which add up to *100*. Given two percent distributions, as ordered lists with *n* components, corresponding to years *s* and *t*, such as $f^s = (f_1^s, f_2^s, \cdots, f_n^s)$ and $f^t = (f_1^t, f_2^t, \cdots, f_n^t)$, the L1 distance between them will be defined as

$$d(f^s, f^t) = (1/2) \sum_{j=1}^{n} \left| f_j^s - f_j^t \right| \ . \tag{3}$$

The factor ½ is used so that the distance takes values between *0* and *100*. This distance will be applied to the original games, as well as to the respective power indices.

4 Results

The results will be illustrated with the example of rice. The annual data, starting in 1975, are available down to the level of municipality. Here, only regional distributions will be shown. It is well known that a motion to the South region has taken place over the years. But a more detailed analysis is required. Table 1 shows a typical situation where power indices can be used to detect changes in the structure of winning coalitions. For any pair of consecutive years, the distance appears on the second line of the pair. For convenience, the distributions corresponding to the original production data are labelled as type "DATA"; the others correspond to the respective power indices.

Table 1 shows rather small distances for the data, and much larger values for the power indices. In particular, from 1989 to 1990, the percent contribution of rice in the South region changed from a value smaller than *50* to a value larger than that. When the data distribution for 1990 is seen as a voting game, it has a dominant player and all the others are dummies. In that kind of situation, the indices of power give a value of *100* to the dominant player and *0* to the others.

Figure 1 illustrates a typical result with regard to the three sets of distances. In several cases, small changes in the distance between the original distributions coincide with abrupt changes in those calculated with the power indices.

Type	Year	North	Northeast	Southeast	South	Center-West	Distance
DATA	1988	9.73	17.62	13.60	40.23	18.82	–
	1989	10.63	16.18	13.18	43.63	16.38	4.30
	1990	8.16	11.53	13.87	54.11	12.33	11.17
BANZHAF	1988	0.00	16.67	16.67	50.00	16.67	–
	1989	9.09	9.09	9.09	63.64	9.09	22.73
	1990	0.00	0.00	0.00	100.00	0.00	36.36
SHAPLEY	1988	0.00	16.67	16.67	50.00	16.67	–
	1989	10.00	10.00	10.00	60.00	10.00	20.00
	1990	0.00	0.00	0.00	100.00	0.00	40.00

Table 1: L1 distances for sets of distributions related to rice production quantity

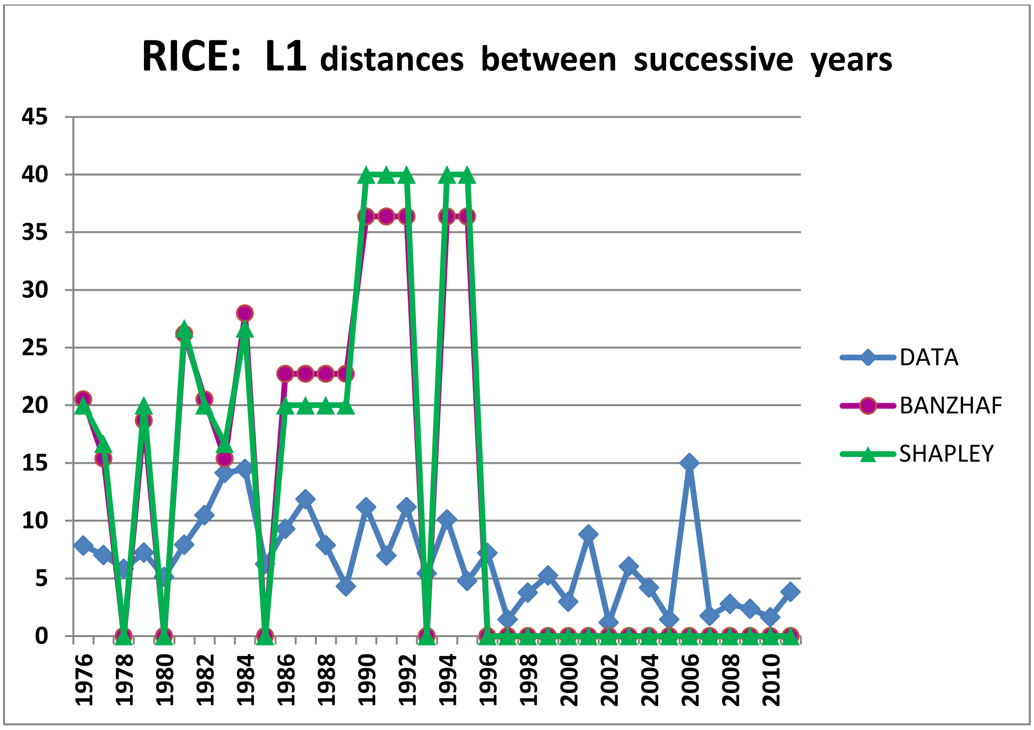

Figure 1: L1 distances, for data and power indices, in successive years

In such cases, the distances between the indices show greater discriminating power than those derived from the original distributions. For that reason, power indices are being used as an auxiliary data mining technique.

References

[1] Dubey, P. and Shapley, L. S. Mathematical properties of the Banzhaf power index. *Mathematics of Operations Research*, 4:99-131, 1979.

[2] Garagorry, F. L. and Chaib Filho, H. Concentration and dynamics of the Brazilian agriculture. In Prado, H. A. do, Luiz, A. J. B. and Chaib Filho, H., editors, *Computational methods for agricultural research: advances and applications*. Hershey, Information Science Reference, 2011.

[3] Owen, G. *Game theory*. New York, Academic Press, 1982. 2nd. ed.