MULTI-CRITERIA ANALYSIS OF BUSINESS INDICATORS BY THE DOMINANCE-BASED ROUGH SET APPROACH

Ayrton Benedito Gaia do Couto
Analista de Sistemas, BNDES
Av. Rep. do Chile, 100, Rio de Janeiro, CEP 20031-917, RJ, Brazil, Phone: +55 21 2172-7658
ayrtoncouto@gmail.com

Luiz Flavio Autran Monteiro Gomes
Professor, Ibmec/RJ
Av. Pres. Wilson, 118, Rio de Janeiro, CEP 20030-020, RJ, Brazil, Phone: +55 21 4503-4053
autran@ibmecrj.br

RESUMO

Este artigo mostra como extrair informações essenciais de um conjunto de indicadores empresariais financeiros e não financeiros, conduzindo a um melhor entendimento do desempenho das empresas. Faz-se uso da Teoria dos Conjuntos Aproximativos e do princípio da Dominância. Tal uso justifica-se pela possibilidade da existência de dados imprecisos, levando à necessidade de tratamento analítico desta imprecisão, de modo a obter-se um conjunto essencial de informações efetivamente consistentes. A análise realizada teve como base a publicação "Exame Melhores e Maiores 2013" (Exame, 2013), das 500 (quinhentas) maiores empresas pelo ordenamento de vendas líquidas, de diferentes setores econômicos. Este estudo sinalizou a importância em se ampliar a análise de indicadores empresariais quando o objetivo é, por exemplo, produzir-se uma visão mais crítica para decisões de investimento.

PALAVRAS CHAVE. Análise Multicritério. Princípio de Dominância. Teoria dos Conjuntos Aproximativos.

Área Principal: DM - Decisão Multicritério.

ABSTRACT

This study demonstrates how to extract essential information from a set of financial and non-financial business indicators, leading to a better understanding of company performance. Rough Set Theory and the Dominance principle are used. This usage is justified by the possibility of there being uncertain data - uncertainty that needs to be treated analytically to yield an essential set of effectively consistent information. The analysis was based on the Brazilian publication Exame Melhores e Maiores 2013, which lists the 500 largest companies in various economic sectors ordered by net sales. This study revealed the importance of broadening the analysis of enterprise indicators when the goal is, for example, to produce a more critical view for investment decisions.

KEYWORDS. Dominance principle. Multi-Criteria Analysis. Rough Set Theory.

Main Area: MD – Multi-criteria Decision.

1.INTRODUCTION

The present study demonstrates how to extract essential information from a set of financial and non-financial business indicators, leading to a better understanding of company performance. It reveals the importance of broadening the analysis of enterprise indicators when the goal is, for example, to produce a more critical view for investment decisions or to identify the performance of probable competitors. The study based on the following research question: "How can patterns be inferred from the analysis of business indicators using a Multi-Criteria approach?". The study is based on a publication listing the 500 largest companies from several sectors of the Brazilian economy (Exame, 2013) and uses Multi-Criteria decision-supporting tools to collectively analyse the various attributes (financial and non-financial indicators) as condition and decision criteria. The choice of Rough Set Theory (RST) and Dominance-based Rough Set Approach (DRSA) as tools to support Multi-Criteria decision is justified by the possibility of there being uncertain (inconsistent) data and the need to address this lack of precision because an information system (or a data table) may need to be processed mathematically. RST was chosen because it does not require any preliminary information about the data at hand, such as its probability distribution. Other theories might be used – e.g., the Fuzzy Set Theory, proposed by Lotfi Asker Zadeh in 1965 (Zadeh, 1965) as an extension of conventional (boolean) logic that introduces the concept of non-absolute truth and serves as a tool for dealing with uncertainties in natural language (Gomes and Gomes, 2012). RST and Fuzzy Set Theory are independent approaches to the treatment of imperfect (incomplete) and uncertain (vague, indeterminate) knowledge (Pawlak et al., 1995). RST has shown to be of fundamental importance to artificial intelligence (AI) and cognitive sciences, especially in the areas of machine learning, knowledge acquisition, decision analysis, knowledge discovery from databases, decision support systems, inductive reasoning and pattern recognition (Pawlak et al., 1995). To support Multi-Criteria analysis, we used the jMAF (Dominance-Based Rough Set Data Analysis Framework) software tool (Blaszczynski et al., 2012), which is available for research purposes by the Institute of Computer Science of the Poznan University of Technology in Poland. The present paper comprises a short introduction to Rough Set Theory (RST) and to the Dominance-based Rough Set Approach (DRSA) in sections 2 and 3, respectively, a study conducted of the 20 (twenty) largest companies based on their net sales and business indicators in section 4, and conclusions and remarks on future studies in section 5.

2.ROUGH SET THEORY

RST had its origin with Zdzislaw Pawlak (Pawlak, 1982): it proposes the treatment of data uncertainty using "lower and upper approximations" for a data set (Pawlak, 1991). One of its concepts, the "indiscernibility relation," identifies objects that

have the same properties, i.e., "indiscernible" objects, to be treated as similar or identical. An information system can be defined as a tuple S = (U, Q, V, f), where U is a finite set of objects, Q is a finite set of attributes, V= $\bigcup_{q \in Q}$ Vq, where Vq is the domain of attribute q, and f: U χ Q $\rightarrow$ V is a total function such that f(x, q) $\in$ Vq for every q$\in$Q, x$\in$U, known as an "information function" (Pawlak and Slowinski, 1994). Given an information system S = (U, Q, V, f), P $\subseteq$ Q, and x,y $\in$ U, we say x and y are "indiscernible" through the set of attributes P in S if f(x,q) = f(y,q) for all q$\in$P. Therefore, all P $\subseteq$ Q generate a binary relation in U, known as an "indiscernibility relation", denoted by IND(P). Given that P $\subseteq$ Q and Y $\subseteq$ U, the lower ($\underline{P}$Y) and upper approximations ($\overline{P}$Y) are defined as:

$$\underline{P}Y = \cup\{X \in U/P : X \subseteq Y\};\ \overline{P}Y = \cup\{X \in U/P : X \cap Y \neq \emptyset\} \quad (1)$$

The difference between $\underline{P}$Y and $\overline{P}$Y is called the "boundary region" of Y:

$$BNP(Y) = \underline{P}Y - \overline{P}Y \quad (2)$$

There is also the concept of accuracy:

$$\alpha P(Y) = \text{card}\ \underline{P}\ /\text{card}\ \overline{P} \quad (3)$$

which captures the degree to which the knowledge of set Y is complete.

There are two more fundamental concepts in RST: an information system's "reduct" and "core". The reduct is its essential part, i.e., the subset of attributes that provides the same quality of classification as the original set of attributes (it allows one to make the same decisions as if all condition attributes were there). The core is the most important subset of this knowledge (Pawlak, 1991; Pawlak and Slowinski, 1994). The family of reducts and the core, should there be any, can also be identified by building a Discernibility Matrix (Chen et al., 2012). As a demonstration of RST's application, refer to the cited example of decision-making in determining the number of executives based on replicated and inconsistent data (Human Resources area) (Couto and Gomes, 2010).

3.DOMINANCE PRINCIPLE

The key aspect of a Multi-Criteria decision is considering objects that are described by multiple criteria and that represent conflicting points of view. Criteria are attributes in domains with an ordering preference; e.g., in choosing a car, one may consider the price and fuel consumption to be characteristics that should serve as criteria in its acquisition, as one usually considers a low price to be better than a high price and moderate fuel consumption to be more desirable than high consumption. In general, other attributes such as colour and country of origin, the domains of which have no ordering preference, are not considered to be decision criteria – they are regular attributes. Therefore, the RST approach does not allow one to analyse multi-criteria decision problems because the analysis uses only regular attributes. Moreover, one cannot identify inconsistencies that violate the following dominance principle: "objects with a better evaluation or having at least the same evaluation (decision

class) cannot be associated to a worse decision class, all decision criteria being considered". RST ignores not only the preference ordering in the set of attributes' values but also the "monotonic" relation of objects' evaluations regarding the condition attributes' values and decision attributes' values' order of preference (classification or degree of preference) (Slowinski et al., 2012; Kotlowski and Slowinski, 2013). This problem is treated in an extension of RST called Dominance-based Rough Set Approach or DRSA (Slowinski et al., 2012), in which indiscernibility relations are replaced with dominance relations in the approximations of decision classes. Furthermore, due to the preferential ordering between decision classes, sets become approximations known as unions of "upward" and "downward" decision classes. Thus, for a tuple S = (U, Q, V, f), set Q is generally divided into condition attributes (set C) and decision attributes (set D). Assuming all condition attributes ($q \in C$) are decision criteria, S_q represents a non-classifiable relation in U with respect to criterion q such that xS_qy denotes "x is at least as good as y in regards to criterion q". Assuming the set of decision attributes D defines a partition of U into a finite number of classes, $Cl = \{Cl_t, t \in T\}$, $T = \{1, ..., n\}$ is a set of these classes such that each $x \in U$ belongs to one and only one $Cl_t \in Cl$. These classes are assumed to be ordered, i.e., for every $r,s \in T$ such that $r > s$, objects of Cl_r are preferable to objects of Cl_s. Therefore, objects can be approximated by unions of "upward" and "downward" decision classes, respectively: $Cl_t^{\geq} = \bigcup_{s \geq t} Cl_s$, $Cl_t^{\leq} = \bigcup_{s \leq t} Cl_s$, t=1, ...,n. The indiscernibility relation is thus substituted with a dominance relation. One says that x dominates y regarding $P \subseteq C$, denoted xD_Py, if xS_qy for all $q \in P$. The dominance relation is reflexive and transitive. Given that $P \subseteq C$ and $x \in U$, the "granules of knowledge" used in the DRSA approximations are:

- a set of dominating objects x, called the P-dominating set: $D_P^+(x) = \{y \in U: yD_Px\}$,
- a set of objects dominated by x, called the P-dominated set: $D_P^-(x) = \{x \in U: xD_Py\}$.

Using the $D_P^+(x)$ sets, the P-lower and P-upper approximations of $Cl_t^{\geq}$ are: $\underline{P}(Cl_t^{\geq}) = \{x \in U: D_P^+(x) \subseteq Cl_t^{\geq}\}$, $\overline{P}(Cl_t^{\geq}) = \bigcup_{x \in Cl_t^{\geq}} D_P^+(x)$, for t=1,...,n. Analogously, the P-lower and P-upper approximations of ($Cl_t^{\leq}$) are: $\underline{P}(Cl_t^{\leq}) = \{x \in U: D_P^-(x) \subseteq Cl_t^{\leq}\}$, $\overline{P}(Cl_t^{\leq}) = \bigcup_{x \in Cl_t^{\leq}} D_P^-(x)$, for t=1,...,n. The P-boundary sets of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ are: $Bn_P(Cl_t^{\geq}) = \overline{P}(Cl_t^{\geq}) - \underline{P}(Cl_t^{\geq})$, $Bn_P(Cl_t^{\leq}) = \overline{P}(Cl_t^{\leq}) - \underline{P}(Cl_t^{\leq})$, , for t=1,...,n. These approximations to the unions of "upward" and "downward" decision classes can be used to infer decision rules of the form "*if ... then ...*". For a given union of "upward" or "downward" of decision classes $Cl_t^{\geq}$ or $Cl_t^{\leq}$, $s,t \in T$, the rules induced under the hypothesis that objects pertaining to lower approximations $\underline{P}(Cl_t^{\geq})$ or $\underline{P}(Cl_t^{\leq})$ are positive and all others are negative suggest that an object be attributed to "at least one class Cl_t" or to "at most one class Cl_s", respectively. These rules are known as

"certain decision rules" ($D_{\leq}$ or $D_{\geq}$) because they attribute objects to unions of decision classes without any ambiguity. Alternatively, if objects pertain to upper approximations, the rules are known as "possible decision rules"; thus, objects could pertain to "at least one class Cl_t" or "at most one class Cl_s". Finally, if objects pertain to the intersection $\overline{P}(Cl_s^{\leq}) \cap \overline{P}(Cl_t^{\geq})$ ($s<t$), the rules induced are known as "approximate rules", i.e., objects are between classes Cl_s and Cl_t. Therefore, if for each criterion $q \in C$, $V_q \subseteq \mathbf{R}$ (V_q is quantitative) and for each $x,y \in$ U, $f(x,q) \geq f(y,q)$ implies xS_qy (V_q has a preferential ordering), decision rules can be considered to be of five types:

1- certain $D_{\geq}$-decision rules:

if $f(x,q_1) \geq r_{q1}$ *and* $f(x,q_2) \geq r_{q2}$ *and ...* $f(x,q_p) \geq r_{qp}$, *then* $x \in Cl_t^{\geq}$

2- possible $D_{\geq}$-decision rules:

if $f(x,q_1) \geq r_{q1}$ *and* $f(x,q_2) \geq r_{q2}$ *and ...* $f(x,q_p) \geq r_{qp}$, *then x possibly belongs to* $Cl_t^{\geq}$

3- certain $D_{\leq}$-decision rules:

if $f(x,q_1) \leq r_{q1}$ *and* $f(x,q_2) \leq r_{q2}$ *and ...* $f(x,q_p) \leq r_{qp}$, *then* $x \in Cl_t^{\leq}$

4- possible $D_{\leq}$-decision rules:

if $f(x,q_1) \leq r_{q1}$ *and* $f(x,q_2) \leq r_{q2}$ *and ...* $f(x,q_p) \leq r_{qp}$, *then x possibly belongs to* $Cl_t^{\leq}$, where $P = \{q_1, ..., q_p\} \subseteq C$, $(r_{q1}, ..., r_{qp}) \in V_{q1} \times V_{q2} \times ... \times V_{qp}$ and $t \in T$;

5- approximate $D_{\leq\geq}$-rules:

if $f(x,q_1) \geq r_{q1}$ *and* $f(x,q_2) \geq r_{q2}$ *and ...* $f(x,q_k) \geq r_{qk}$ *and* $f(x,q_{k+1}) \leq r_{qk+1}$ *and* $f(x,q_p) \leq r_{qp}$, *then* $x \in Cl_s \cup Cl_{s+1} \cup ... \cup Cl_t$.

Rules of types "1" and "3" represent "certain knowledge" extracted from a data table (or information system), rules of types "2" and "4" represent "possible knowledge", and the rule of type "5" represent "ambiguous knowledge". Some of its basic concepts are as follows: a rule's "strength" is the ratio of the number of objects that satisfy the rule to the total number of objects, its "certainty" is the ratio of the number of objects that satisfy the rule to the number of objects that satisfy the rule's condition criteria, and its "coverage" is the ratio of the number of objects that satisfy the rule to the number of objects that satisfy the rule's decision criteria.

4. MULTI-CRITERIA ANALYSIS OF BUSINESS INDICATORS

According to (Pace et al., 2003), intangible assets as a decisive factor in obtaining competitive advantage characterises became increasingly important at the end of the 20th century. Organisations look for ways to better measure and present these intangible assets to managers and investors. Aggregate measures, such as ROI (Return on Investment), ROE (Return on Equity) and operating profit, are no longer able to capture the complexity and values in the business environment. This environment is now process oriented, and the predominant aspects are related to identifying opportunities, speed of learning, innovation, quality, flexibility, reliability and response capacity, all of which must be measured. Analysts who use this type of non-financial information are capable of producing the most accurate projections; if they can communicate these attributes, they will more easily obtain funding from third

parties. Studies show that investors want to understand the business model in greater depth, with a view into the main performance indicators, and they pay attention to non-financial aspects, which they use in guiding their decisions to invest. Table 1 shows these measures organised into nine categories:

(A) Financial
(B) Product Quality
(C) Client Satisfaction
(D) Process Efficiency
(E) Product and Process Innovation
(F) Competitive Environment
(G) Management Quality and Independence
(H) Human Resources Administration
(I) Social Responsibility

A - Financial	
1-	Net profit and profit/share
2-	Cash flow
3-	ROE (Return on Equity)
4-	ROA (Return on Assets)
5-	Sales
6-	Return on Sales
7-	Sales/Total Assets
8-	Net Equity / Total Assets
9-	Quality of accounting practices
B - Product Quality	
10-	% repeated sales
11-	Clients that improve company image
12-	Complaints within warranty
13-	Client complaints
C - Client Satisfaction	
14-	Market research
15-	Punctual delivery
16-	Service response time
17-	% returning customers
18-	NAV (net asset value)
19-	% contacted clients who effectively buy
20-	Litigation with clients
D - Process Efficiency	
21-	Rate of defects
22-	Product development time
23-	Manufacturing cycle time
24-	Time between ordering and delivery
25-	Capacity for customization
26-	Operational costs /employee
27-	Sales/employee
28-	COGS (Cost of goods sold)/ stock
29-	Accounts receivable/sales
30-	Capital investment
31-	Plant/equipment age
32-	Usage of installed capacity
E - Product/Process Innovation	
33-	R&D expenditures
34-	% of patented products
35-	Number of new patents
36-	Number of new products
37-	% of new product sales
F - Competitive Environment	
38-	Market share
39-	Brand recognition
40-	Potential competition
41-	Tax/quota protection
42-	% sales of patented products
43-	Strategic alliances
44-	Litigation w/ anti-trust legislation
45-	Geographic diversification
46-	Client diversification
47-	Product diversification
G - Management Quality/Independence	
48-	Management continuity
49-	Managers' experience/reputation
50-	Managing council's involvement
51-	Managing council's independence
52-	Litigation with shareholders
53-	Weakening of control
54-	Managers' ethical behaviour
55-	Value offered to investors
H - Human Resources Administration	
56-	Equal opportunity employment
57-	Employee involvement
58-	Profit sharing
59-	Stock options plan
60-	% candidates to positions in competing companies successfully recruited
61-	Job/employee development
62-	% new employees
63-	Benefit policy
I - Social Responsibility	
64-	Minority protection
65-	Performance in environmental activities
66-	Involvement with communities
67-	Litigation

Table 1
Measures under study; source: (Pace et al., 2003)

Of these measures, those were sought that would generate the most aggregated value, as shown in Table 2.

Measure	Ct	M	Measure	Ct	M
Clients that improve company image	B	4.00	Litigation w/ anti-trust legislation	F	2.64
Weakening of control	G	4.00	Profit sharing	H	2.62
Managing council's independence	G	4.00	Managing council's involvement	G	2.62
Sales	A	3.90	Time between ordering and delivery	D	2.49
Cash flow	A	3.70	R&D expenditures	E	2.45
Return on Sales	A	3.40	% sales of patented products	F	2.44
Usage of installed capacity	D	3.35	% returning customers	C	2.36
Market share	F	3.31	% of new product sales	E	2.31
Net profit and profit/share	A	3.30	Stock options plan	H	2.31
Client diversification	F	3.29	Performance in environmental activities	I	2.31
Accounts receivable/sales	D	3.29	% repeated sales	B	2.31
Capital investment	D	3.27	Number of new products	E	2.24
Potential competition	F	3.27	Market research	C	2.20
Geographic diversification	F	3.24	% contacted clients who effectively buy	C	2.16
Plant/equipment age	D	3.22	Capacity for customization	D	2.16
ROE (Return on Equity)	A	3.20	Client complaints	B	2.11
Product diversification	F	3.18	Service response time	C	2.10
Sales/ total assets	A	3.10	Rate of defects' amount	D	2.07
Net Equity / total Assets	A	3.10	Product development time	D	2.07
Quality of accounting practices	A	3.10	Litigation	I	2.07
Sales/employee	D	3.09	Minority protection	I	2.04
Operational/employee costs	D	3.05	Job/employee development	H	2.00
Tax/quota protection	F	3.04	Punctual delivery	C	2.00
Litigation with shareholders	G	3.00	% of patented products	E	1.98
ROA (Return on Assets)	A	3.00	Number of new patents	E	1.98
Management continuity	G	2.98	Litigation with clients	C	1.96
COGS (Cost of goods sold)/ stock	D	2.93	Involvement with communities	I	1.96
Managers' experience/reputation	G	2.89	Benefit policy	H	1.96
Value offered to investors	G	2.82	Complaints within warranty	B	1.95
Brand recognition	F	2.80	Employee involvement	H	1.93
Strategic alliances	F	2.76	Equal opportunity employment	H	1.90
Managers' ethical behaviour	G	2.75	% candidates to positions in competing companies successfully recruited	H	1.85
Manufacturing cycle time	D	2.73	% new employees	H	1.75
NAV (net asset value)	C	2.71			

Table 2
Capacity to predict value (measure, category, mean score)
source: (Pace et al., 2003)

The top 20 companies (for which all information needed for this study was available) were selected from among the 500 largest companies in various economy sectors from the net sales ranking published in Exame magazine (Exame, 2013). They are shown in Table 3, where their names have been replaced with letters in the same order as their net sales ranking. The productivity metric "revenue created per employee" was unavailable for some companies; in these cases, the "net sales per employee" metric was adopted, considering "productivity" to be the ratio of output(s) to input(s) (Bandeira, 2009).

Based on Table 2, we sought to select the information (measures) that aggregates the most value to a company. For example, "Market share" (mean 3.31) is preferable to "R&D expenditures" (mean 2.45). Market share can be considered to be an indicator of "efficacy" as the ratio of actual to expected outputs – in this case, (company's net sales / total net sales of the economy sector) x 100 (Bandeira, 2009). This group of the 20 largest companies has representatives from the following sectors of the economy: wholesale, automotive industry, consumer goods, energy, mining, chemical and petrochemical, services, steelworks and metallurgy, telecommunications and retail.

Company	Ranking	Net Sales (US$ million)(1)	Legal Net Profit (US$ million)(2)	Legal Profitability (%) (3)	Net working capital (US$ million) (4)	General liquidity (index number) (5)	General debt (%) (6)	EBITDA (US$ million) (7)	Net sales/employee (US$ million) (8)	Market share (%) (9)
A	1	109,713.30	10,225.10	5.9	13,801.30	0.57	38.9	22,774.70	1.8	46.5
B	2	39,024.50	925.40	17.1	1,217.50	1.77	39.8	1,588.90	8.7	30.7
C	3	28,989.40	4,763.40	6.0	5,096.20	0.45	36.5	16,348.40	0.6	70.4
D	4	23,596.60	379.70	27.0	852.20	0.74	71.9	685.60	10.6	18.6
E	5	11,914.90	197.80	17.8	1,353.80	0.89	71.8	761.70	1.7	10.2
F	6	11,708.80	590.10	31.8	270.80	0.76	85.8	1,265.90	0.6	12.6
G	7	11,484.40	2,042.60	32.4	814.60	0.92	55.9	3,876.50	0.9	16.8
H	8	11,099.40	8.10	0.2	3,439.50	1.60	48.9	724.20	1.7	9.5
I	9	10,416.00	-357.80	-8.0	-780.20	0.43	74.6	1,138.30	2.1	17.2
J	10	9,617.20	514.40	12.0	-640.10	0.62	61.6	756.10	0.2	9.3
K	11	7,193.80	398.00	5.5	388.70	0.78	52.2	306.80	0.1	6.2
L	12	7,052.10	510.90	21.1	241.80	1.42	60.2	-33.70	0.1	8.1
M	13	6,584.70	5,142.20	29.7	-2,237.70	0.46	41.4	1,972.60	0.3	5.7
N	14	6,503.40	2,179.40	9.3	296.10	0.77	22.6	1,962.80	1.2	9.5
O	15	6,448.60	-468.60	-6.9	-325.30	0.46	50.0	511.80	0.8	12.8
P	16	5,761.80	-313.00	-3.8	557.40	0.58	43.6	37.60	0.4	11.4
Q	17	5,420.20	935.60	15.3	-209.20	0.28	54.5	1,767.30	0.4	6.3
R	18	5,371.20	-205.60	-4.6	1,314.30	0.31	80.9	1,403.70	0.3	10.6
S	19	5,164.10	341.50	10.2	1,684.90	1.06	60.2	820.00	0.3	5.6
T	20	5,027.30	52.80	2.2	250.10	0.67	59.8	1,242.40	0.9	2.1

(1) value of gross sales, from which returns, discounts and sales taxes were deducted
(2) nominal result for the period (not considering inflation), deducting income taxes and social contribution and adjusting interest on net equity
(3) main indicator of business excellence (return on investment): = profit (net profit, legal, adjusted/net equity, legal, adjusted) x 100
(4) short-term resources available for financing company's activities
(5) = (working assets + long-term receivables) / total enforceable; less than 1 implies solvency will depend on future profits, debt negotiation or sale of assets.
(6) business risk; = [(working liabilities + non-working liabilities)/adjusted total assets] x 100
(7) cash flow generated by business activity (profit before deducting interest, taxes on profits, depreciation and amortisation)
(8) net sales/mean number of employees; productivity indicator
(9) (net sales/total net sales in the economy sector) x 100

Table 3
The 20 largest companies by net sales; source: (Exame, 2013)

The data in Table 3 were then analysed using the jMAF software application (Blaszczynski et al., 2012). Three analyses were conducted:

a) Analysis 1: All attributes were considered according to their nature, which we denote by *gain* when increasing values mean greater advantage, *cost* when increasing values mean smaller advantage, and *none* otherwise, i.e.,:

**ATTRIBUTES
+ company: (nominal), none, description
- ranking: (integer), none
+ net_sale: (continuous), gain, decision
+ net_profit: (continuous), gain
+ profitability: (continuous), gain
+ net_ working _capital: (continuous), gain
+ equity: (continuous), gain
+ debt: (continuous), cost
+ ebitda([1]): (continuous), gain
+ sale_per_employee: (continuous), gain
+ market_share: (continuous), gain
decision: net_sale

([1]) earnings before interest, taxes, depreciation and amortization

As an example, the (condition) criterion "debt" was categorised as "cost", i.e., it is expected that a company's results will be better when its debt is smaller. Because the criterion chosen to order the largest companies was "net sales", it was used as a "decision" criterion in this analysis. This kept a "1:1" relationship between the order and net sales. For classification, we used DRSA with the VC-DRSA method, which permitted reclassifying objects that violated the Dominance principle (companies F and G). This analysis is shown in Table 4.

Company	Ranking	Net Sales (US$ million)	Legal Net Profit (US$ million)	Legal Profitability (%)	Net working capital (US$ million)	General liquidity (index number)	General debt (%)	EBITDA (US$ million)	Net sales/employee (US$ million)	Market share (%)
A	1	109,713.30	10,225.10	5.9	13,801.30	0.57	38.9	22,774.70	1.8	46.5
B	2	39,024.50	925.40	17.1	1,217.50	1.77	39.8	1,588.90	8.7	30.7
C	3	28,989.40	4,763.40	6.0	5,096.20	0.45	36.5	16,348.40	0.6	70.4
D	4	23,596.60	379.70	27.0	852.20	0.74	71.9	685.60	10.6	18.6
E	5	11,914.90	197.80	17.8	1,353.80	0.89	71.8	761.70	1.7	10.2
F	6	11,708.80	590.10	31.8	270.80	0.76	85.8	1,265.90	0.6	12.6
G	7	11,484.40	2,042.60	32.4	814.60	0.92	55.9	3,876.50	0.9	16.8
H	8	11,099.40	8.10	0.2	3,439.50	1.60	48.9	724.20	1.7	9.5
I	9	10,416.00	-357.80	-8.0	-780.20	0.43	74.6	1,138.30	2.1	17.2
J	10	9,617.20	514.40	12.0	-640.10	0.62	61.6	756.10	0.2	9.3
K	11	7,193.80	398.00	5.5	388.70	0.78	52.2	306.80	0.1	6.2
L	12	7,052.10	510.90	21.1	241.80	1.42	60.2	-33.70	0.1	8.1
M	13	6,584.70	5,142.20	29.7	-2,237.70	0.46	41.4	1,972.60	0.3	5.7
N	14	6,503.40	2,179.40	9.3	296.10	0.77	22.6	1,962.80	1.2	9.5
O	15	6,448.60	-468.60	-6.9	-325.30	0.46	50.0	511.80	0.8	12.8
P	16	5,761.80	-313.00	-3.8	557.40	0.58	43.6	37.60	0.4	11.4
Q	17	5,420.20	935.60	15.3	-209.20	0.28	54.5	1,767.30	0.4	6.3
R	18	5,371.20	-205.60	-4.6	1,314.30	0.31	80.9	1,403.70	0.3	10.6
S	19	5,164.10	341.50	10.2	1,684.90	1.06	60.2	820.00	0.3	5.6
T	20	5,027.30	52.80	2.2	250.10	0.67	59.8	1,242.40	0.9	2.1

Table 4
Companies F and G violate the Dominance principle regarding the class “net sales”

Considering that companies are ordered by net sales and that company G is superior to company F in all indicators, then by the Dominance principle, company G should be placed in the same class of net sales as company F, or in some class above. As a result of this analysis, 49 new rules were inferred.

For example, for the rule “3” that was generated, “strength” and “coverage” values were computed for each inferred rule (“bold-face emphasis ours”):

rule 3: (market_share>= 30.7) => (net_sale >= 28989.4) |CERTAIN, AT_LEAST, 28989.4|
LearningPositiveExamples: 1, 2, 3
Support: 3
SupportingExamples: 1, 2, 3
Strength: 0.15
Confidence: 1.0
CoverageFactor: 1.0
Coverage: 3
CoveredExamples: 1, 2, 3

From rule "3" the following is inferred: "for those companies whose market share was greater than or equal to 30.7% (companies A, B and C; supporting examples: 1, 2, 3), net sales was greater than or equal to US$ 28,989.40 million". This means 15% (strength = 3/20 or 0.15) of the 20 largest companies satisfy this rule, and 100% (coveragefactor = 3/3 or 1.0) of the companies with net sales equal to or greater than US$ 28,989.40 million also satisfy this rule.

b) Analysis 2: In this case, following a suggestion given by jMAF, the following “reducts” were chosen: {profitability, equity, ebitda, sales per employee, market share}. There was again an indication of reclassification of companies F and G as to the ordering by net sales, as shown in Table 4. For example, for generated rule “22”, the following levels of “strength” and “coverage” of each inferred rule were computed (the bold-faced emphasis is ours):

rule 22: (profitability>= 5.5) & (market_share >= 6.2) => (net_sale >= 5420.2) |CERTAIN, AT_LEAST, 5420.2|
LearningPositiveExamples: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17
Support: 12
SupportingExamples: 1, 2, 3, 4, 5, 6, 7, 10, 11, 12, 14, 17
Strength: 0.6
Confidence: 1.0
CoverageFactor: 0.7058823529411765
Coverage: 12
CoveredExamples: 1, 2, 3, 4, 5, 6, 7, 10, 11, 12, 14, 17

The "22" previous rule inferred that "companies with higher profitability or equal 5.5% and market share in excess or equal to 6.2%, then its net sales were greater than or equal to US$ 5,420.20 million". This means that 60% (strength = 12/20 or 0.6) of the 20 largest companies satisfy this rule, and 70.5% (coverage factor = 12/17 or 0.705) of the companies with net sales equal to or greater than US$ 5,420.20 million also satisfy this rule.

c) Analysis 3: For this analysis, as suggested by jMAF, "profitability" was chosen as the "core" of the condition attributes or criteria. Taking only the "profitability" condition criterion into account, the following 6 rules were inferred.

[RULES]
#Certain at least rules
1: (profitability >= 31.8) => (net_sale >= 11484.4) |CERTAIN, AT_LEAST, 11484.4|
2: (profitability >= 17.1) => (net_sale >= 6584.7) |CERTAIN, AT_LEAST, 6584.7|
3: (profitability >= 12.0) => (net_sale >= 5420.2) |CERTAIN, AT_LEAST, 5420.2|
4: (profitability >= 5.5) => (net_sale >= 5164.1) |CERTAIN, AT_LEAST, 5164.1|
#Certain at most rules
5: (profitability <= -3.8) => (net_sale <= 10416.0) |CERTAIN, AT_MOST, 10416.0|
6: (profitability <= 5.5) => (net_sale <= 11099.4) |CERTAIN, AT_MOST, 11099.4|

The levels of "strength" and "coverage" were computed for rule "4" (as before, the bold-faced emphasis is ours):

rule 4: (profitability >= 5.5) => (net_sale >= 5164.1) |CERTAIN, AT_LEAST, 5164.1|
LearningPositiveExamples: 1, 2, 3, 4, 5, 6, 7, 10, 11, 12, 13, 14, 17, 19
Support: 14
SupportingExamples: 1, 2, 3, 4, 5, 6, 7, 10, 11, 12, 13, 14, 17, 19
Strength: 0.7
Confidence: 1.0
CoverageFactor: 0.7368421052631579
Coverage: 14
CoveredExamples: 1, 2, 3, 4, 5, 6, 7, 10, 11, 12, 13, 14, 17, 19

From the above, the following can be inferred: "for companies with higher profitability than or equal to 5.5%, its net sales were greater than or equal to US$ 5,164.10 million".

5. CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

It is hard to obtain companies' non-financial indicators because, traditionally, only a company's financial health is analysed. However, in the face of the need to establish long-term strategies, surveying non-financial indicators may signal where it is possible to aggregate value to the company's business relative to changes in the competition, markets and technology.

The present study shows that the analysis can be extended by taking other financial and non-financial indicators into account to infer patterns that are often hidden in the data. Despite the relatively small set of 20 largest companies in a universe of the 500 largest by net sales, this study showed the importance of broadening the analysis of business indicators when the goals are, for example, to produce a more critical view for investment decisions, to identify the performance of probable competitors and to ascertain market trends. As the study was based on a Multi-Criteria analysis, Rough Set Theory and the Dominance principle were used to identify and treat data that, in some cases, proved to be inconsistent or to create paradoxes when objects' classifications did not fit into some reference standard ("decision class") – e.g., the case of companies F and G, see Table 4. In this case, for an investment decision in a company, for example, company F would most likely not have so much of an advantage (in net sales) relative to company G. Thus, the analysis of these 20 largest companies using the jMAF software application (Blaszczynski et al., 2012) allowed us to infer hidden patterns when (1) all financial and non-financial indicators were taken into account as condition criteria (except for "net sales", taken as a decision criterion) – Analysis 1; (2) when a given set of condition attributes was selected as the "reduct", i.e., equivalent to the set of all attributes – Analysis 2; and (3) when the condition attribute or criterion "profitability" was used as the "core" of the set of attributes – Analysis 3. In all these analyses, the attribute "net sales" was kept as the decision criterion.

Regarding the ordering of companies by net sales as suggested in the cited study, it is possible to address this in future studies with the analysis results shown in Table 3 by processing these with the jRank application (*Ranking using Dominance-based Rough Set Approach*) (Szelag et al., 2013), which also applies concepts of Rough Set Theory and the Dominance principle to establish a classification order. Furthermore, it is also possible to perform a broader analysis with the entire 500 largest companies to infer patterns of these companies' financial and non-financial indicators.

ACKNOWLEDGMENT

The work leading up to this paper was partially supported by CNPq/Brazil, through Research Project n. 305732/2012-9. The authors thank the Institute of Computer Science of the Poznan University of Technology, Poland, for allowing them to use the jMAF software in their research.

REFERENCES

Bandeira, A.A. (2009), Indicadores de desempenho: instrumentos à produtividade organizacional [Performance indicators: instruments for organisational productivity], Rio de Janeiro: Qualitymark.

Blaszczynski, J., Greco, S., Matarazzo, B., Slowinski, R. and Szeląg, M. (2012), jMAF, Institute of Computer Science. Poznan University of Technology, Poland. Available from: http://www.cs.put.poznan.pl/jblaszczynski/Site/jRS.html.

Chen, D., Zhao, S., Zhang, L., Yang, Y. and Zhang, X. (2012), Sample Pair Selection for Attribute Reduction with Rough Set, IEEE Transactions on Knowledge and Data Engineering, 24, 11, 2080-2093.

Couto, A.B.G. and Gomes, L.F.A.M. (2010), A tomada de decisão em recursos humanos com dados replicados e inconsistentes: uma aplicação da teoria dos conjuntos aproximativos [Decision-making in human resources with replicated and inconsistent data: an application of rough set theory], Pesquisa Operacional, 30, 3, 657-686.

Exame (2013), Exame Melhores e Maiores [Exame Best and Biggest],São Paulo: Abril, 1044 E. ed.

Gomes, L.F.A.M. and Gomes, C.F.S. (2012), Tomada de decisão gerencial: enfoque multicritério [Management decision-making: multi-criterion approach], São Paulo: Atlas, 4. ed.

Kotlowski, W. and Slowinski, R. (2013), On nonparametric ordinal classification with monotonicity constraints, IEEE Transactions on Knowledge and Data Engineering, 25, 11, 2576-2589.

Pace, E.S.U., Basso, L.F.C. and Silva, M.A. (2003), Indicadores de Desempenho como Direcionadores de Valor, [Performance Indicators as Value Drivers], RAC, 7, 1, 37-65.

Pawlak, Z. (1982), Rough sets, Int. J. Comput. Inf. Sci. 11, 341-356.

Pawlak, Z. (1991), Rough sets. Theoretical aspects of reasoning about data, Kluwer Academic Publishers, Dordrecht.

Pawlak, Z., Grzymala-Busse, J., Slowinski, R. and Ziarko, W. (1995), Rough sets, Communications of the ACM, 38, 11, 89-95.

Pawlak, Z. and Slowinski, R. (1994), Rough set approach to multi-attribute decision analysis, European Journal of Operational Research, Invited Review, 72, 443-459.

Slowinski, R., Greco, S. and Matarazzo, B. (2012), Rough set and rule-based multicriteria decision aiding, Pesquisa Operacional, 32, 2, 213-269.

Szelag, M., Slowinski, R., Greco, S., Blaszczynski, J. and Wilk, S. (2013), jRank, Institute of Computer Science, Poznan University of Technology, Poland. Available from: http://www.cs.put.poznan.pl/mszelag/Software/jRank/jRank.html.

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