Accounting Information Inconsistencies and their Effects on Insolvency Prediction Models

Ricardo Lopes Cardoso\textsuperscript{a,b,c}, Alexandre Mendes\textsuperscript{d}, Poueri do Carmo Mario\textsuperscript{e}, Antonio Lopo Martinez\textsuperscript{f} and Felipe Ramos Ferreira\textsuperscript{b,g}

\textsuperscript{a} International Accounting Standards Committee Foundation, Education Projects, London, United Kingdom.
\textsuperscript{b} The Brazilian School of Public and Business Administration, Getulio Vargas Foundation, Rio de Janeiro, RJ, Brazil.
\textsuperscript{c} Business and Finance School, State University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil.
\textsuperscript{d} School of Electrical Engineering and Computer Science, The University of Newcastle, Callaghan, NSW, Australia.
\textsuperscript{e} School of Economics, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil.
\textsuperscript{f} Business School, FUCAPE, Vitoria, ES, Brazil.
\textsuperscript{g} Accounting School, Centro Universitario do Para, Belem, PA, Brazil.

Abstract

Many studies have shown that avoiding political costs is an incentive for firms to manipulate accounting information, e.g., McNichols and Wilson (1988); Jones (1991); Kato et al. (2001). The majority of them use discretionary accruals models as proxies to manipulation. This paper introduces a new variable (DIF) that measures data inconsistencies present in financial reports, replacing discretionary accruals in the detection of manipulation. Accounting information was collected from 2,033 Brazilian health maintenance organizations (HMOs). This information was then processed and financial ratios derived from insolvency prediction models and thresholds established by the regulatory agency were taken as attributes to differentiate solvent HMOs from insolvents. During this data processing, inconsistencies were identified and instead of being removed, were used to determine the value of the attribute DIF. Processed data was then analysed using data mining techniques and a series of classifiers were created. The classifiers found have high accuracy in terms of discriminating distressed companies, especially when the DIF variable is used. In addition, the attributes selected and the structure of the classifiers can be supported by traditional models of analysis based on financial ratios. Results are relevant for those who are interested in assessing firms’ insolvency risk, because data inconsistencies may signal firms’ performance, therefore shall not be removed from analysis.

Keywords: insolvency; inconsistency; earnings management; health maintenance organization

JEL descriptors:
M410 – Accounting
G330 – Bankruptcy; Liquidation
G220 – Insurance: Insurance Companies
C810 – Methodology for Collecting, Estimating, and Organizing Microeconomic Data

1. Introduction

This paper addresses two important topics in the analysis of insolvency prediction for health maintenance organizations (hereafter, HMO). First is the discussion of insolvency and prediction models specifically for HMOs. The second and most relevant topic refers to the manipulation of accounting data by organizations and how that might mislead regulators while assessing firms solvency and affect prediction models. However, before we address these themes in more depth, it is necessary to provide some background information about insolvency and bankruptcy, and the scenario of HMOs in Brazil.
Insolvency and bankruptcy have been studied in the areas of Accounting and Finance for several decades. Most of these studies address these elements under different points of view; either trying to predict them (Altman, 1968; Ohlson, 1980; Newton, 2003; Altman and Hotchiss, 2006), or analysing the processes that occur during an insolvency crisis or bankruptcy (Aghion et al., 1992, 1993; Hart, 1997, 2000). In the literature, insolvency, failure and bankruptcy usually appear as synonyms; however they refer to different moments. Insolvency is linked to a state; failure to an act; and bankruptcy has a legal meaning, as in a judicial process.

By definition, insolvency is defined as the state in which the company is not capable of honouring some commitment. The Brazilian law follows the same definition but specifically mentions defaulted payment as the main characteristic to determine insolvency (or distress).

The HMO industry in Brazil is subjected to a specific regulation and the phenomenon of bankruptcy of Brazilian HMOs can be divided into two phases. The first continues until the HMO becomes insolvent. The second phase starts when the Brazilian federal regulatory agency (the Brazilian Agência Nacional de Saúde Suplementar, hereafter ANS) – equivalent to the U.S. Federal Government’s Office of Health Maintenance Organizations – becomes aware of the insolvency situation and starts administrative processes that might culminate in the declaration of bankruptcy. Between those two phases there is a time lag. Typically, the time lag between the company becoming insolvent and the declaration of bankruptcy corresponds to the time between the failure to honour a payment, and the creditor’s start of a bankruptcy procedure. For Brazilian HMOs, this procedure might be started by any stakeholder making a specific claim or by the routine auditing conducted by the ANS, based on quarterly financial reports.

The detection of insolvency via the auditing process conducted by the ANS is very complex and fuzzy; although it is based on financial ratio parameters (see Table 1 for details). Since 2001, the ANS has adopted three levels of insolvency risk: low, medium, and high. Accounting information is analysed every quarter, if the HMO does not break any threshold established by the ANS, it is considered to be low risk and continues to operate normally. If the HMO breaks some of those thresholds, its risk is classified as medium. The HMO would then go through continuous, more rigorous analysis and could be asked to provide a so-called ‘recovery plan’. The recovery plan is subject to approval by the ANS and consists of a monthly budget based on a set of operating, investing and financing decisions in order to minimize the HMO’s insolvency risk. In this case the company is also obliged to present financial data in a monthly basis to the ANS and to fulfil other documents. Finally, if the HMO breaks the majority of thresholds established by the ANS and/or its recovery plan is not successful, its insolvency risk is re-classified as high. The HMO then becomes subject to a direct intervention by the ANS, which may lead to the organization’s discontinuity and assets’ liquidation. Several parameters are used to identify the financial state of an HMO, including current ratio, profitability, return on assets, and return on equity, among others, as presented in Table 1.

This scenario creates the opportunity for companies in financial distress to manipulate their accounting information, as discussed in Benham (2005) – response to regulation, and in Laughlin (2007) – regulation of accounting. In fact, we suspect that HMOs in financial distress manipulate their accounting information in order to introduce noise and mislead the process conducted by the ANS to evaluate their solvency status.

Earnings management literature, mainly based on discretionary accruals as proxy for manipulation, has shown that banks (Kato et al, 2001; Fuji, 2004), HMOs (Mensah et al., 1994; Cardoso, 2005), and producers in general (Jones, 1991; Navissi, 1999) manipulate their accounting information in order to mislead regulators. Differently from previous studies, this study is not based on any discretionary accruals models. Instead, it relies on the consistency of HMO’s quarterly accounting data.

To the best of our knowledge, this study is the first attempt to consider a variable that captures accounting data inconsistence on insolvency prediction models. There are some similarities between this study and Cormier et al. (2000) and Kasanen et al. (1996). However, compared to those studies, this study’s relative contributions can be summarized as follows. First, those studies focus on the accounting relevance for the capital market, providing evidences from the correlation between accounting data and price of securities; while the present study focuses on the accounting relevance for assessing financial
distress. Second, Cormier et al. (2000) and Kasanen et al. (1996) use discretionary accruals as proxy for unreliable accounting information, while we used a case-based investigation on quarterly financial reports on the search for inconsistencies (the parameters we used to assess inconsistency are listed in the subsection 3.1). A part from these two most relevant differences, while those studies investigate Swiss and Finnish listed companies from many different economic sectors, this paper investigates Brazilian health maintenance organizations (HMOs) which are private firms (except by eight HMOs that either are listed companies or their parent are – notice that the entire sample comprises 2,033 HMOs).

2. Theoretical background

This section presents a brief summary about the manipulation of accounting information (Section 2.1) and about insolvency prediction models (Section 2.2).

2.1 Manipulation of accounting information

According to the definitions of Schipper (1989), Healy and Wahlen (1999), Fields et al. (2001) and Mckee (2005), the manipulation of accounting information is the choice of accounting practices or operational decisions with the goal to elaborate reports and report financial numbers different from those that would be presented if such practices were not adopted. Therefore, the goal – to portrait a specific financial position and performance – can be achieved through accounting practices and operating decisions.

Accounting decisions involve the choice of accounting practices related to: (a) identification of the phenomenon – transactions and other events; (b) measurement of their effect on the firm’s performance and net assets; (c) classification; (d) accounting recognition; and (e) presentation and disclosure of the firm’s financial position. In the literature there are numerous examples of manipulation of accounting information through misleading accounting practices. Among them we must cite McNichols and Wilson (1988); Jones (1991); Dechow et al. (1995); and Kang and Sivaramakrishnan (1995).

In Brazil, the study of Martinez (2001) based on a sample of non-financial companies traded at the local stock market has shown that the most common manipulation of accounting information aims to avoid the reduction of the net profit, as well as to reduce its volatility (also referred to as income smoothing). Also, Fuji (2004) has shown that in a sample composed of the 50 largest Brazilian banks, the manipulation of accounting information is concentrated on the use of the provision account for allowance for bad debts. It aims at reducing the political cost related to the regulation made by the Brazilian Central Bank. There are several other examples in the same line. Cardoso, 2005, based on a sample composed of quarterly financial information from 2001 through 2003 of more than 1,000 HMOs, has shown that HMOs tend to manipulate accounting information in order to avoid breaking financial thresholds established by the ANS (specifically to avoid reporting net loss and negative owners’ equity).

A second type of manipulation of accounting information is done using operating decisions. Mckee (2005) explains this type of manipulation with an example which mentions the implementation (or not) of special discounts or special programs to increase sales close to the end of a quarter in which the income goals were not achieved. Other types of operating decisions include the investment in new equipment, hiring of new staff, etc. These types of manipulation impact the company’s cash flow and consequently the income and expenditures associated to these activities. There are very few studies in the international literature that deals with this kind of manipulation (Roychowdhury, 2003, 2005; Gunny, 2005; Zang, 2005). Based on Brazilian companies, Martinez and Cardoso (2009), using a sample composed of non-financial, non-insurance and non-HMO firms have shown that companies that operate on a more restrictive regulatory environment tend to prefer the manipulation via operating decisions than accounting practices.

Notwithstanding these two ways of analysing the manipulation of accounting information, we did not estimate manipulation from any econometric models presented on both literature. Instead, we used a case-based evidence of manipulation, i.e. data inconsistencies, as described in Section 3.
Data inconsistencies are analysed here as the regulation of accounting (Laughlin, 2007) – or a response to regulation (Benham, 2005). HMOs manipulate accounting information, through a simplistic way (i.e., providing inconsistent data to ANS), because the ANS regulates the accounting policies HMOs might adopt and uses accounting information to assess whether an HMO is solvent or insolvent.

Such regulation is based on ‘market-wide cost savings from regulation’ logic. Where the costs of complying with a one-size-fits-all regime are relatively low, standardization of corporate reporting can make it easier for the agency to process the information and to compare across companies (Leuz and Wysocki, 2008).

Considering the way accounting policies are regulates and accounting information used by the ANS, and that information is provided in an environment characterized by uncertainty and imperfect information, which leads to asymmetric knowledge and transaction costs, HMOs’ managers have different incentives with regards to accounting information from stakeholders (including ANS’s staff). Therefore, we suspect that HMOs in financial distress manipulate their accounting information in order to introduce noise and mislead the process conducted by the ANS to evaluate their solvency status.

2.2 Insolvency prediction models

The literature on insolvency prediction models is absolutely vast. For a comprehensive literature review, we suggest Altman and Narayanan (1997), Altman and Saunders (1998), and Balcaen and Ooghe (2004). Following, we present a brief comment about the classical references on this issue, namely: Altman (1968) and Ohlson (1980).

The most relevant contribution of Altman (1968) was the usage of multivariate data analysis technology to predict the insolvency of firms. On a sample comprised by 66 firms, with half on them going bankrupt in the period 1946–1965, Altman (1968) used discriminant analysis to identify which financial ratios could better discriminate insolvent firms from solvent ones. Altman (1968) ended up with a model (Z-Score) composed of five accounting based ratios:

\[
Z\text{-Score} = 0.012x_1 + 0.014x_2 + 0.033x_3 + 0.006x_4 + 0.999x_5, \text{ where:}^1
\]

- \(x_1\) = Common size ratio for Working Capital
- \(x_2\) = Common size ratio for Retained Earnings
- \(x_3\) = Operating Cash Flow to Asset ratio
- \(x_4\) = Equity to Debt ratio
- \(x_5\) = Asset’s turnover

The classification of firms into solvent and insolvent was based on the mean of the median results for each cluster of firms. But the author did not consider the costs associated with misclassification or errors probability. Altman’s (1968) main results are: (i) the power of accounting based ratios to predict insolvency gets impaired on the years close to the bankruptcy itself and (ii) the most relevant changes on financial ratios occur in between two and three years prior to the bankruptcy.

Twelve years later, Ohlson (1980) presented an insolvency prediction model which main improvement, compared to the predecessors models, was the usage of another statistic technique, namely the conditional logit. This technique is less dependent on assumptions than the discriminant analysis and provides a probability function as result.

Based on a sample of 105 listed companies that went bankrupt in the period 1970–1976 and of 2,058 solvent listed companies randomly selected, Ohlson (1980) built a model comprised of nine financial ratios: \(X_1 = \text{SIZE}\); \(X_2 = \text{TLTA}\); \(X_3 = \text{WCTA}\); \(X_4 = \text{CLCA}\); \(X_5 = \text{OENEG}\); \(X_6 = \text{NITA}\); \(X_7 = \text{FUTL}\); \(X_8 = \text{INTWO}\); \(X_9 = \text{CHIN}\).

1 See table 1, for details about each indicators.
Ohlson’s (1980) main results are: (i) the size of the entity is relevant to the classification and (ii) increasing the gap between the date when accounting information was disclosed and the date when the entity was declared bankrupted negatively affects the power of the model.

Although both literature tracks are mature, there is a gap on the link between them. Therefore, the research question addressed in this paper is: Are insolvency prediction models improved by the use of an attribute related to the reliability of accounting information? The logic underling this research question relies on the literature of regulation and earnings management. Managers of regulated firms have incentives to create noise on information used by the regulatory body in order to keep information asymmetry and mislead regulators’ decisions (notwithstanding the market-wide cost savings that justify this type of regulation). Thus, it is expected that an attribute based on data inconsistencies should be relevant to insolvency prediction models.

3. Data collection and processing

The data used in this study was obtained directly from the ANS in 2008\(^2\). The accounting information comprises 2,033 HMOs and contains all their financial reports between 2001 and 2007, plus the first two quarters of 2008. It contains some inconsistencies, missing information, among other problems; therefore, some data processing was necessary before the analysis could be carried out.

3.1 Data processing

The first need for adjustments identified refers to changes in parameters. In that case, adjustments had to be made to keep consistency along the years. As an example, the balance from actuarial contingencies liability had to be added to current liabilities, as from 2001 to 2006 they were accounted for separately, in conformity to the official accounts plan, whereas from 2007 onwards they were merged. Such technical changes are not the focus of this study, so will not be described but only stated that adjustments were carried out.

However, one type of adjustment is worth mentioning and requires a thorough description. While processing the data, we detected many differences (or errors) on the HMOs ending balances; which we simply called ‘inconsistencies’. Instead of treating them as missing values, we decided to keep them in our dataset by incorporating a categorical variable named DIF, which assumes values 0 or 1. Zero means that less than 50% of the quarterly information presents accounting inconsistencies for the HMO; and 1, that 50% or more of the quarterly information presents inconsistencies. The criteria to identify an inconsistency are:

- Any of the following accounts has a negative balance: Current Assets, Non-Current Assets, Monetary Current Assets, Prepaid Commercial Expenses, Current Liabilities, Non-Current Liabilities, Monetary Current Liabilities, or Contributed Capital;
- Any of the following accounts has a balance lower than R$ 1,000.00\(^3\): Total Assets, Total Liabilities, or Total Revenues;
- Total Assets differs from the sum of Total Liabilities and Owners’ Equity;
- Total Assets differs from the sum of Current Assets and Non-Current Assets;
- Total Liabilities differs from the sum of Current Liabilities and Non-Current Liabilities;
- Net Income differs from the sum of Total Revenues and Total Expenses;
- Total Assets balance is lower than Owners’ Equity;
- The variable Size (measured in accordance to Table 1) has a negative value;
- Any of the following variables is a missing value: Size or Cost of Debt.

Another important point is that, in order to use data mining techniques we had to combine the yearly

\(^2\) Data was obtained through an agreement with the ANS and the public institution that funded this research (CNPq), process MCT/CNPq/ANS 410612/2006-5.

\(^3\) R$ 1,000.00 = CAD$ 577.90 = US$ 561.04 = €412.11, as of March 5th, 2010. Hence, any of those accounts presenting such a low balance was considered an inconsistency.
information contained in the dataset. That is, for each attribute and for each company, we had several
data points, each corresponding to an accounting period, between 2001 and 2008. This information had
to be combined into a single value, which would characterize that attribute for that company. In our
study, after testing both median and mean, we opted for the median, as it was less affected by the
presence of outliers.

3.2 Data attributes and parameters

Three datasets were generated for this study. The first dataset is named $\text{DIF}_0$ and contains 1,239
companies; all with $\text{DIF} = 0$ (i.e., HMOs which less than 50% of the quarterly information presents
accounting inconsistencies). The second is named $\text{DIF}_{01}$ and contains all the 2,033 companies, but the
same attributes as in the dataset $\text{DIF}_0$ (i.e., does not include the attribute $\text{DIF}$). Finally, the third dataset
is named $\text{DIF}_{01}^*$, and contains all the 2,033 companies, as well, but the attribute $\text{DIF}$ is included in the
analysis. Apart from $\text{DIF}$, the same set of attributes was considered for analysis in the three datasets.
These attributes, as well as a brief explanation of their meaning and theoretical source are shown in
Table 1.

4. Methods

In this section we will describe the data mining technique$^4$ used to generate the classifiers to determine
insolvency for the companies that compose the three datasets: $\text{DIF}_0$, $\text{DIF}_{01}$, and $\text{DIF}_{01}^*$.

Data mining is the process to extract patterns in large amounts of data. The technique is applied to
several areas in which the detection of patterns, or profiles, is necessary; e.g. marketing, surveillance,
fraud detection, as well as other areas of science. In the past, this kind of information used to be
extracted manually. However, with the increase in data availability due to the introduction of computers
in businesses, this task quickly became too time-consuming. Automated techniques had to be developed,
which resulted in the four main branches of data mining research:

- **Classification** - Organizes objects, or information, into pre-defined groups.
- **Clustering** - Similar to classification, but the groups are not pre-defined; therefore, the method will
group objects with similar characteristics, without a priori information about the groups.
- **Regression** - Involves the determination of a function that models the data with the least error.
- **Association rules** - Involves the search for relations between the attributes present in the data.

This work focuses on the first branch, i.e. classification. In this case, the goal is to determine the most
relevant attributes to classify companies as solvent/insolvent with the greatest accuracy; and generate
classification models. To characterize insolvency, we have used the attribute ‘Insolvent’ (see Table 1),
which indicates whether or not the company is inactive. If the company is inactive, it is an indication of
insolvency, and this attribute assumes a value 1. If the value of Insolvent is 0, then the company is still
active. Hereon, we will use the term active as a synonym to solvent; analogously to inactive/insolvent.

The active/inactive variable is the status as how each HMO appears on ANS’s database. If active, the
firm still operates. If inactive, the firm ceased its operations.

This work focuses on the first branch, i.e. classification. In this case, the goal is to determine the most
relevant attributes to classify companies as solvent/insolvent with the greatest accuracy; and generate
classification models. To characterize insolvency, we have used the attribute ‘Insolvent’ (see Table 1),
which indicates whether or not the company is inactive. If the company is inactive, it is an indication of
insolvency, and this attribute assumes a value 1. If the value of Insolvent is 0, then the company is still
active. Hereon, we will use the term active as a synonym to solvent; analogously to inactive/insolvent.

---

$^4$ We don’t explain about Discriminant Analysis and Logit Model, used in Altman (1968) e Ohlson (1980) in this
study. We understood that techniques were explored in many other articles and now, is not more difficult learn
about it.

---
Table 1: List of attributes considered for analysis in datasets DIF0, DIF01 and DIF01*, divided according to their source.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Attribute Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cds</td>
<td>Cost of Debt IE/(STD+LTD)</td>
</tr>
<tr>
<td>Csc</td>
<td>Liability Structure CL/TL</td>
</tr>
<tr>
<td>COMBa</td>
<td>Expense Index (MCE+CE+AE)/OR</td>
</tr>
<tr>
<td>COMDBa</td>
<td>Amortized Expense Index (MCE+CE+AE)/(OR-ID+II)</td>
</tr>
<tr>
<td>CT/CP</td>
<td>Debt to Equity ratio TL/OE</td>
</tr>
<tr>
<td>Daa</td>
<td>Admin. Expenses to Revenue ratio AE/OR</td>
</tr>
<tr>
<td>Dox</td>
<td>Commerical Expenses to Revenue ratio CE/OR</td>
</tr>
<tr>
<td>Dmo</td>
<td>Med Care Expenses to Revenue ratio MCE/OR</td>
</tr>
<tr>
<td>ENDIV or TLTA</td>
<td>Common Size ratio for Debt TL/TA</td>
</tr>
<tr>
<td>GAFs</td>
<td>Financial Leverage RE/RA</td>
</tr>
<tr>
<td>GAoa</td>
<td>Operating Leverage (ROI1-ROI2)/(ROI1-ROI2)</td>
</tr>
<tr>
<td>GATa</td>
<td>Return on Asset's Turnover TA/OR</td>
</tr>
<tr>
<td>LNOG</td>
<td>Working Capital Need Index (CCE-STD)/((CA-CL)+(CCE-STD))</td>
</tr>
<tr>
<td>IMOBREC</td>
<td>Common Size ratio for Long-term Assets LTA/PA</td>
</tr>
<tr>
<td>LC</td>
<td>Current Ratio CA/CL</td>
</tr>
<tr>
<td>LG</td>
<td>Expanded Current Ratio (CA+LIR)/(CL4-NCL)</td>
</tr>
<tr>
<td>NCOA</td>
<td>Common Size ratio for Working Capital Need (CA-CL)-(CCE-STD)/TA</td>
</tr>
<tr>
<td>NCGRa</td>
<td>Working Capital Need to Revenue ratio (CA-CL)+(CCE-STD))/OR</td>
</tr>
<tr>
<td>PMCr</td>
<td>Receivables conversion ratio 360x:AR/OR</td>
</tr>
<tr>
<td>PMPhi</td>
<td>Payable conversion ratio 360x:AP/MCE</td>
</tr>
<tr>
<td>Rass</td>
<td>Operating Income to Asset ratio OI/TA</td>
</tr>
<tr>
<td>RPEta</td>
<td>Return on Equity (ROE) NI/OE</td>
</tr>
<tr>
<td>TA</td>
<td>Log of Total Asset log(TA)</td>
</tr>
<tr>
<td>Trn</td>
<td>Treasury to Revenue ratio (CCE-STD)/OR</td>
</tr>
<tr>
<td>VARLL</td>
<td>Net Income Variation (NIIQ1-NIIQ4)/NIIQ4</td>
</tr>
<tr>
<td>VARREC</td>
<td>Revenue Variation (ORQ1-ORQ4)/ORQ4</td>
</tr>
<tr>
<td>WCLA</td>
<td>see OCLA</td>
</tr>
</tbody>
</table>

Source: Altman, 1968

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Attribute Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GIB0Aa</td>
<td>Asset's Turner OR/TA</td>
</tr>
<tr>
<td>GLCAa</td>
<td>Operating Cash Flow to Asset ratio EBIT/TA</td>
</tr>
<tr>
<td>LAA</td>
<td>Common Size ratio for Retained Earnings (OE-CC)/TA</td>
</tr>
<tr>
<td>FLOT</td>
<td>Equity to Debt ratio OE/TL</td>
</tr>
<tr>
<td>CCLA or WCTA</td>
<td>Common Size ratio for Working Capital (CA-CL)/TA</td>
</tr>
</tbody>
</table>

Source: Ohlson, 1980

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Attribute Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCLA or WCTA</td>
<td>Common Size ratio for Working Capital (CA-CL)/TA</td>
</tr>
<tr>
<td>CHINa</td>
<td>Net Income Performance (NIIQ1-NIIQ4)/(NIIQ1-NIIQ4)</td>
</tr>
<tr>
<td>CIRC</td>
<td>Revenue of Current Ratio CL/CA</td>
</tr>
<tr>
<td>FULLa</td>
<td>Operating Cash Flow to Debt ratio EBIT/(TA-OE)</td>
</tr>
<tr>
<td>INTWOa</td>
<td>2 years of losses 1 if there were Net Losses in the past two years; 0 otherwise.</td>
</tr>
<tr>
<td>NIHa</td>
<td>Net Income to Asset NI/TA</td>
</tr>
<tr>
<td>OGENEG</td>
<td>Negative Owner's Equity 1 if there is Negative Owner's Equity; 0 otherwise.</td>
</tr>
<tr>
<td>SIZE</td>
<td>Size log(TA)/GPIC</td>
</tr>
<tr>
<td>TLTa</td>
<td>see ENDIV</td>
</tr>
</tbody>
</table>

New variable introduced in this study

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Attribute Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIF</td>
<td>Inconsistency 0 if less than 95% of the quarters have inconsistencies; 1 otherwise.</td>
</tr>
</tbody>
</table>

The active/inactive variable is the status as how each HMO appears on ANS’s database. If active, the firm still operates. If inactive, the firm ceased its operations.

Although there are other reasons for a HMO to become inactive (e.g. a HMO can be merged to another, sell its clients portfolio, or discontinue its operations even though being capable of paying its obligations), and it is not necessarily true that every active HMO is solvent, we consider inactivity as the best proxy for measuring insolvency for the following reasons:
Data accessibility - The database provided by the ANS does not provide the HMOs’ name, but only fictitious nicknames (i.e. the names are coded). Hence, the use of information from credit agencies about the HMOs’ payment default is not possible.

Political choice - ANS’s decisions concerning to risk classification (low, medium or high) and their consequent administrative processes (recovery plan or intervention), are probably biased by political issues. Hence, using them as proxy to insolvency would create an internal validity problem in our research. In this case, we would identify which attributes lead the ANS to open an administrative process (presumably based on insolvency signs) instead of identifying which attributes discriminate solvent HMOs from insolvent ones, regardless of the ANS’s decisions.

The type of classifier chosen for this work is based on decision trees. In its graph structure the leaves of the tree represent the final classification – in our case, either active or inactive – and the internal nodes and branches represent the attributes and the thresholds that lead to the classification.

There are several algorithms available that generate decision trees. Among the most traditional ones, we must cite the ID3 (Iterative Dichotomiser 3), introduced by Quinlan, 1986. Also, important is the C4.5 (Quinlan, 1993), which is an extension of the ID3 that uses the concept of ‘information entropy’ to determine the most relevant attributes to the classification. It also works with missing values and discrete variables. Finally, there is the J48 (Witten and Frank, 2005), which is an extension of the C4.5, with parameters to adjust the size of the classification tree and the minimum number of elements that are classified in each leaf, among others.

In this study, we used the J48 algorithm, which is available in the data mining package WEKA\(^5\), from the University of Waikato, New Zealand. It is an open source, public domain software for data analysis. With its first version distributed in 1997, WEKA became a reference tool for data mining, with over 8,000 citations in the scientific literature.

The algorithm behind the J48 is described next:

1. Starting with the root node, check among all attributes which one discriminates the examples present in the dataset (i.e. find the attribute that generates the highest information gain). That will create the first division, composed of the root node and two branches.

2. For each new division created, verify whether or not all examples that are directed to a branch belong to the same class. If that happens, then that branch does not need to be divided any further and is declared a leaf node, whose class is determined by the class of the examples.

3. If the examples that follow a given branch belong to different classes, find the attribute that generates the highest information gain for those examples. If there is no increase in the information gain, then that leaf node is declared as being of the same class as the majority of the examples. Otherwise, a new division is created.

4. The process is repeated until all branches end in leaf nodes.

Next, we will describe the computational tests, as well as the parameters used to produce them and the results.

5. Results

In this section, we compare the use of the DIF attribute and decision trees against a traditional insolvency models techniques. Two traditional insolvency models were used: Altman (1968) and Ohlson (1980). Results for all three techniques are presented for three dataset (DIF\(_0\), DIF\(_{01}\) and DIF\(_{01*}\)).

5.1 Results from the proposed models using data mining and the attribute DIF

The tests were divided into three parts, one for each dataset – DIF\(_0\), DIF\(_{01}\) and DIF\(_{01*}\). For each subsection, given that the number of active companies is considerably higher than inactive ones, our initial tests would always generate classifiers which would be biased in favour of the dominant class. To

\(^5\) http://www.cs.waikato.ac.nz/ml/weka
reduce this tendency, we instructed WEKA to always select a proportion of 1:1 of active/inactive companies before creating the models. This has reduced the bias and generated models that could classify better both active and inactive companies.

The decision tree models were created using all examples in each case, that is, training and test sets were the same. The parameters of the model were kept as the default suggested in WEKA. The only changes were in ‘reduced error pruning’, which was set to true; and the minimum number of objects in each leaf ($\text{minNumObj}$) which was set at 5% of the number of active and inactive companies in each dataset. Therefore, for $\text{DIF}_0$, $\text{minNumObj} = 5$; and for $\text{DIF}_0^*$ and $\text{DIF}_0^{**}$, $\text{minNumObj} = 23$. These parameters were manually set to obtain a well-balanced trade-off between the complexity of the decision tree, i.e. the number of attributes used, and the accuracy of the classification.

The results for data mining presented in the following subsections show that the inclusion of the attribute DIF in the tests improved the accuracy of and simplified the insolvency prediction model, being the attribute DIF the first node in the decision tree presented in Figure 3.

### 5.1.1 Results for $\text{DIF}_0$ – data mining

This group contains 1,239 companies – 101 inactive and 1,138 active. Since WEKA selected examples in a proportion of 1:1, all 101 inactive companies were included, and a random batch of 101 active companies, out of 1,138 was also selected. The resulting decision tree, as well as a table with the statistics of the classification is shown in Figure 1. The model correctly classified 64.8% of the examples. The attributes chosen were $\text{LC}$, $\text{VAR_LL}$ and $\text{GATa}$. The numbers in the leaf nodes indicate the class (0 = active; 1 = inactive). Also, in parentheses, the first number indicates the number of examples classified in that node, and the second number indicates the number of wrongly classified examples. The non-integer values are the result of a stochastic procedure from J48 to deal with attributes with missing values, which are present in the datasets.

The result is in agreement with the accounting theory. If the current ratio is low ($\text{LC} \leq 0.363$), the HMO is immediately identified as inactive (insolvent). If that is not the case, but the net profit has suffered a large reduction ($\text{VAR_LL} \leq -1.963$) and the inverse of the assets’ turnover is high ($\text{GATa} > 0.338$), the HMO is also classified as inactive (insolvent).

![Figure 1: Decision tree and classification statistics for the health maintenance organizations that present inconsistencies in less than 50% of their quarterly information (i.e. DIF$_0$).](image)

### 5.1.2 Results for $\text{DIF}_0^*$ – data mining

This group contains 2,033 companies – 466 inactive and 1,567 active. Again, because of the 1:1 proportion of active/inactive companies, all 466 inactive companies were used in the construction of the model, plus a random sampling of 466 active companies. The model is showed in Figure 2. The accuracy of the model comprising all entities is higher than the accuracy of the model that used only a subsample (entities that presented less than 50% of quarterly information with inconsistencies, $\text{DIF}_0$) – 69.1% compared to 64.8%. This increase may be a consequence of the significative increase in the number of entities analysed (202 firms in $\text{DIF}_0$ dataset compared to 932 firms in the $\text{DIF}_0^*$ dataset – an
increase of 361.4%).

In addition, the resulting model is considerably more complex – the decision tree has five internal nodes with attributes, compared to only three for DIF$_0$. The attributes selected from DIF$_{01}$ are all different from those selected from DIF$_0$. The model DIF$_{01}$ selected the attributes VAR_REC, LG, NCGRA and SIZE. An interesting point is that VAR_REC was used twice, indicating that values either too low or too high are indicative of insolvency. The relation low-VAR_REC and insolvency is more intuitive and in agreement with the literature, but the opposite is not. However, high values of revenue variation might be indicative of assets liquidation to improve the financial situation on the surface, but which eventually leads to bankruptcy. In this case, companies with small, positive revenue variations appear to be financially healthier, on average. All other attributes and thresholds present in the decision tree are in agreement with the literature.

Notice that this dataset comprises HMOs that provided reliable and unreliable information. Such complexity of the model may be explained by the noise created by unreliable information.

Figure 2: Decision tree and classification statistics for all health maintenance organizations. In this model, we have the introduction of the DIF attribute, which differentiates companies that presented inconsistencies in the majority of their quarterly information, from those that did not. DIF was positioned at root node of the tree, as the most relevant attribute.

5.1.3 Results for DIF$_{01*}$ – data mining

This group contains the same 2,033 companies used in the dataset DIF$_{01}$, but the attribute DIF was included in the analysis. The model shown in Figure 3 correctly classified around 3/4 of the examples. This indicates that the use of the DIF attribute improved the accuracy by 4.8 percentage points – from 69.1% (dataset DIF$_{01}$) to 73.9% (dataset DIF$_{01*}$). This indicates that if the attribute DIF is ignored, the model’s prediction power is negatively impacted.

The attributes chosen were DIF, VAR_REC and LG. All attributes and thresholds present in the decision tree are in agreement with the literature. For instance, if the accounting information is not reliable (DIF > 0) and the revenue has suffered a large reduction (VAR_REC <= -0.055) then the HMO is classified as inactive (insolvent). If the revenue has not suffered such reduction (but information is still not reliable) and the expanded current ratio is low (LG <= 0.136) then the HMO is also classified as inactive/insolvent.

Notice that the consideration of the attribute DIF, that measures the reliability of the accounting information, improved the model in two perspectives. First, the attribute DIF increased the accuracy of the model. Second, it made the model more simple and easy to understand; it captures the noise created by unreliable accounting information.
Figure 3: Decision tree and classification statistics for all health maintenance organizations, without considering the attribute DIF in the analysis. This absence caused a worsening in the accuracy of the model, which was reduced from 74.2% to 69.1%. In addition, the model became considerably more complex, with five internal nodes, compared to only three for the model Figure 2. This is a clear indicative of the importance of the attribute DIF for the classification of solvency status.

5.2 Results from Altman’s (1968) model

With Altman’s model, the goal is to evaluate the quality of accounting ratios with a structured statistical technique, which uses multivariate discriminate analysis. In order to compare the approaches, the sample of HMOs was separated into active and inactive categories and split between training and test subsamples.

As like as for the tests presented under section 5.1, the tests based on the model developed by Altman (1968) were also divided into three parts, one for each dataset – DIF0, DIF01 and DIF1. For each subsection, given that the number of active companies is considerably higher than inactive ones, we used the same samples that were used in the data mining tests presented in section 5.1.

The results presented bellow in this section show that the Altman’s model only worked reasonably well for the dataset that ignores HMOs that provided more than 50% of quarterly financial reports with at least one of the inconsistencies listed in section 3.1 (i.e., DIF0 dataset). When the entire sample was analysed, the original Altman’s model and the modified-Altman model (with the variable DIF) misclassified more than 70% of inactive HMOs as active. Therefore, we are not confident on those models. However, this fact reinforces the relevance of assessing the reliability of accounting information before analysing firms’ financial performance and financial position.

5.2.1 Results for DIF0 – Altman (1968)

This group contains 1,239 companies – 101 inactive and 1,138 active. We used the same companies used for the data mining test presented in subsection 5.1.1. Among all 202 HMOs for DIF0, the system randomly selected 124 for the training subsample and used the remaining cases for testing. The model correctly classified 62.8% of the testing subsample.

Table 2 presents the confusion matrix and the classification statistics. For active entities, the regression function Z-Score0 is:

\[
Z-\text{Score}_0 = -1.692 + 0.838WCTA + 0.02GLCAa + 0.021PLCT + 0.637Giroa - 0.007LAA
\]
and for inactive HMOs, the regression function $Z_{\text{Score}_1}$ determined is:

$$Z_{\text{Score}_1} = -1.472 + 0.022WCTA + 0.041GLCAa + 0.038PLCT + 0.506GIROa - 0.033LAA$$

Note that the signal for LAA is different from that on the original model. The model misclassified 51.2% of the inactive HMOs as active.

Table 2: Confusion matrix and classification statistics for the health maintenance organizations that present inconsistencies in less than 50% of their quarterly information (i.e. $DIF_{01}$), based on the Altman’s model.

<table>
<thead>
<tr>
<th>Confusion matrix (Training data)*</th>
<th>Confusion matrix (Cross-validated)**</th>
<th>Confusion matrix (Testing data)***</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>Inactive</td>
<td>← Classified as</td>
</tr>
<tr>
<td>49</td>
<td>15</td>
<td>Active</td>
</tr>
<tr>
<td>48</td>
<td>16</td>
<td>Active</td>
</tr>
</tbody>
</table>

*66.1% of selected original grouped cases correctly classified. 124 cases included in analysis.

**62.9% of selected cross-validated grouped cases correctly classified. 124 cases included in analysis.

***62.8% of unselected original grouped cases correctly classified. 78 cases included in analysis.

5.2.2 Results for $DIF_{01}$ – Altman (1968)

This group contains 2,033 companies – 466 inactive and 1,567 active. We used the same companies used for the data mining test presented in subsection 5.1.2. Among all 932 HMOs for $DIF_{01}$, there are some missing values for variable relevant to the Altman’s model (205 HMOs were excluded from analysis). For all other cases, the system randomly selected 416 for the training subsample and used the remaining cases for testing. The model correctly classified 60.8% of the testing subsample. For active entities, the regression function is:

$$Z_{\text{Score}_0} = -1.393 + 0.493WCTA - 0.111GLCAa + 0.01PLCT + 0.693GIROa - 0.033LAA$$

and for inactive HMOs, the regression function $Z_{\text{Score}_1}$ determined is:

$$Z_{\text{Score}_1} = -1.464 + 0.418WCTA - 0.116GLCAa + 0.021PLCT + 0.527GIROa - 0.207LAA$$

Note that the signals for GLCAa and LAA are different that the original model. The model misclassified 83.9% of the inactive HMOs as active, that is very high for a prediction model. Therefore, the results are not presented in a table.

The worsening of the model (compared to the one presented in section 5.2.1) may be caused by the inclusion of data from HMOs that provided more than 50% of quarterly information with at least one of the inconsistencies listed in section 3.1 (unreliable data), created noise in the dataset. Once such noise was not controlled (the variable DIF was not included in this model), the model is not able to properly discriminate active from inactive HMOs.

5.2.3 Results for $DIF_{01*}$ – Altman (1968)

This group contains 2,033 companies – 466 inactive and 1,567 active. We used the same companies
used for the data mining test presented in subsection 5.1.3 and included the variable DIF to the original explanatory variables from Altman’s model. Among all 932 HMOs for DIF01, there are some missing values for variable relevant to the Altman’s model and or on the variable DIF (217 HMOs were excluded from analysis). For all other cases, the system randomly selected 410 for the training subsample and used the remaining cases for testing. Therefore, this section presents the result for a modified-Altman model.

The model correctly classified approximately 61.4% of the testing subsample. For active entities, the regression function is:

\[
Z_{\text{Score}}_0 = -6.701 + 0.411WCTA - 0.143GLCAa + 0.009PLCT + 0.893GIROa - 0.023LAA + 11.765DIF
\]

and for inactive HMOs, the regression function \(Z_{\text{Score}}_1\) determined is:

\[
Z_{\text{Score}}_1 = -8.155 + 0.326WCTA - 0.153GLCAa + 0.019PLCT + 0.751GIROa - 0.195LAA + 13.211DIF
\]

Note that the signal for LAA is different from that on the original model (similar to the dataset DIF0, see section 5.2.1). The model misclassified more than 71.3% of the inactive HMOs as active. Although this result is not as bad as that from the test where the variable DIF was omitted, it is also a very high error level for a prediction model (similar result to that obtained with the dataset DIF01, see section 5.2.2). Therefore, the results are not presented in a table.

Such improvement is a consequence of the inclusion of the variable DIF, that controls the noise created by unreliable data, but the discriminant analysis technique was not able to properly use the variable DIF (structured in this paper as a dichotomy variable).

5.3 Results from the traditional models – Ohlson (1980)

This model uses another statistical technique, namely \textit{logit} (logistic conditional regression), as briefly commented in section 2.2.

As like as for the tests presented under sections 5.1 and 5.2, the tests based on the model provided by Ohlson (1980) were also divided into three parts, one for each dataset – DIF0, DIF01 and DIF01*. For each subsection, given that the number of active companies is considerably higher than inactive ones, we used the same samples that were used in the data mining tests.

The results presented in the subsections that follow show that the original Ohlson’s model did not perform well neither with the dataset DIF0 nor with the dataset DIF01. They misclassified more than 68% of the inactive HMOs as active. However, when the variable DIF was included in the tests (modified-Ohlson model – DIF01* dataset), such error dropped to 48%, being the variable DIF statistically significant at 0.01. Indeed, this evidence reinforces the need to assess the reliability of accounting information before analysing the firms’ financial position and financial performance. Furthermore, it also reinforces the results presented in the subsection 5.1.3 (DIF01* dataset – data mining), that the attribute DIF is relevant for classifying active (solvent) and inactive (insolvent) HMOs.

5.3.1 Results for DIF0 – Ohlson (1980)

This group contains 1,239 companies – 101 inactive and 1,138 active. We used the same companies used for the data mining test presented in subsections 5.1.1 and 5.2.1. Only two variables would be included in the model at the significance level of 0.1, they are: SIZE and CLCA. The model misclassified 70.7% of the inactive HMOs as active. Therefore the results are not reliable and a table is not presented.
5.3.2 Results for DIF$_{01}$ – Ohlson (1980)

This group contains 2,033 companies – 466 inactive and 1,567 active. We used the same companies used for the data mining test presented in subsections 5.1.2 and 5.2.2. However, because of missing values for some relevant variables used in the Ohlson’s model, in this test only 780 HMOs were used (not 932 as in the data mining – as presented in section 4, the J48 algorithm available in the data mining package WEKA is not constrained by missing values). The classification of the HMO’s had an accuracy of 60.7% for the testing subsample.

The variables used by this model were SIZE, INTWOa, WCTA and TLTA, the first three variables at the significance level of 0.01 and the last one at 0.05.

Similarly to the results presented in section 5.3.1, the misclassifications of inactive HMOs as active is very high (68%), indicating that the model is prone to classify organizations with high risk of insolvency as low risk. Therefore, the results are not reliable and a table is not presented.

5.3.3 Results for DIF$_{01*}$ – Ohlson (1980)

The sample is the same used for the data mining test presented in subsection 5.1.3 and included the variable DIF to the original explanatory variables from Ohlson’s model. Therefore, this section presents the result for a modified-Ohlson model. However, because of missing values for some relevant variables used in the Ohlson’s model or on the variable DIF, in this test only 769 HMOs were used. The resulting confusion matrix, as well as the statistics of the classification is shown in Table 3. The classification of the HMO’s had an accuracy of 62.7% for the testing subsample.

The variables used by this model are DIF, SIZE, INTWOa and WCTA, the first three variables at the significance level of 0.01 and the last one at 0.05.

The misclassifications of inactive HMOs as active is reduced from 68% to 48.3%, showing that the inclusion of the variable DIF in the model was able to capture the noise created by unreliable accounting information and is a relevant attribute to discriminate solvent from insolvent HMOs. That is a similar result to that obtained from the comparison of the decision trees for the datasets DIF$_{01}$ and DIF$_{01*}$ (see Section 5.1.2 and 5.1.3, respectively).

Table 3: Confusion matrix and classification statistics for the health maintenance organizations. In this model, we have the introduction of the DIF attribute, which differentiates companies that presented inconsistencies in the majority of their quarterly information, from those that did not. Modified-Ohlson model.

<table>
<thead>
<tr>
<th>Confusion matrix (Training data)*</th>
<th>Confusion matrix (Testing data)**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>Inactive</td>
</tr>
<tr>
<td>195</td>
<td>59</td>
</tr>
<tr>
<td>82</td>
<td>107</td>
</tr>
</tbody>
</table>

*68.1% of selected original grouped cases correctly classified. 442 cases included in analysis.

**62.7% of unselected original grouped cases correctly classified. 396 cases included in analysis.

As presented above, considering all relevant factors, none of the traditional (and original) models is appropriate. They are both significantly vulnerable to Type I Error (classifying inactive entities as active). It also suggests that the variables used in those traditional models do not have a reasonable predictive power to classify Brazilian HMOs between solvent (active) and insolvent (inactive).
6. Discussion

The first finding is that neither traditional model (i.e., Altman, 1968 and Ohlson, 1980) provided an acceptable level of misclassification (i.e., both incurred in significative levels of Type I error). The only exception is in regard with the modified-Ohlson model (with the variable DIF).

The second relevant finding is that none of the explanatory variables provided by traditional insolvency prediction models (e.g. Altman, 1968; Ohlson, 1980) was selected by the data-mining technique to discriminate solvent (active) from insolvent (inactive) HMOs. However, for the dataset composed only by data with DIF = 0 (dataset DIF_0, presented on Section 5.1.1), the data-mining technique selected two attributes (LC, current ratio; and GATa, the reverse of assets’ turnover). Those are the inverse of explanatory variables presented on those traditional models: LC = 1/CLCA from Ohlson, 1980, and GATa = 1/GIROa from Altman, 1968 (see Table 1). Data-mining also selected attributes that the ANS considers relevant for its continuous monitoring activity. This first finding suggests that HMOs have particular features, so may not be analysed as ordinary commercial entities, in the lights of the original insolvency prediction models.

The third relevant finding is that data inconsistency can be seen as a signal that the HMO is facing financial problems. When all HMOs were analysed (dataset DIF_01, presented in Section 5.1.3), the variable inconsistency (DIF) was selected by the data-mining technique as the first attribute relevant to discriminate solvent and insolvent HMOs. On the other hand, when the same set of organizations was analysed but the attribute DIF was ignored, the model’s accuracy dropped significantly, from 74.2% to 69.1%. At the same time, the complexity of the model increased (section 5.1.2). A similar evidence is provided in section 5.3.3, where is shown that the accuracy of the modified-Ohlson model (with the variable DIF) is improved in comparison with the original model. This finding needs to be carefully understood: it does not mean that inconsistency on financial reports leads entities to insolvency (as one might understand by comparing the results from the datasets DIF_01 and DIF_01*). However, it means that HMOs facing financial distress intend to mislead their regulators, analysts and stakeholders in general, or do not have enough infrastructure or knowledge to present financial reports properly.

There are many sophisticated ways to mislead financial information (see McKee, 2005; Palepu and Healy, 2006; Mulford and Comiskey, 2005). Presenting inconsistent financial reports is perhaps the most simplistic one. The use of sophisticated accounting choices requires advanced accounting knowledge, a know-how that Brazilian HMOs do not seem to have. The Brazilian HMO industry is characterized by small and medium sized entities. Only eight (Amil, Dix, Medial Saúde, Odontoprev, Gama Saúde, Gama Odonto, Odonto Empresa and Previdonto) in 2,033 HMOs have their shares (or their parental entity’s shares) publicly traded at the Brazilian Stock and Exchange house – BM&FBovespa; and only 80 in 2,033 HMOs are classified by the ANS as large-sized HMOs, i.e., manage more than 100,000 lives.

Linking this finding to the reviewed literature, results show that firms will respond to regulation when there is a significant constrain imposed by the regulator (Laughlin, 2007; Benham, 2005). In this particular case, the ANS requires all HMOs to send their quarterly financial reports in XML format. If an HMO does not make its quarterly financial report available to the ANS in a defined due date, it is fined by the regulatory agency. However, if the HMO presents any financial information in its quarterly financial report in time, even inconsistent, that information is accepted by the ANS and a fine is not issued. Then, if inconsistencies are detected, the ANS requires the HMO to correct and re-send the information, but still there is no punishment (except when the ANS’s staff realizes that the HMO’s manager was not acting in good-faith).

Finally, the financial reports are used by the ANS to monitor the HMOs’ financial performance and insolvency risk. Depending on financial ratios results (either: low insolvency risk; medium risk or high risk) the ANS decides which actions it will take (either: do nothing; ask for a recovery plan; or start a direct intervention, respectively). Through a transaction cost perspective, it is cheaper for an HMO to present some information (even inconsistent), than not presenting any financial reports at all. This choice avoids immediate monetary punishment (the fine) and delays real insolvency risk detection.
Considering that primarily HMOs facing financial distress have incentives to present inconsistent and misleading financial reports (Akerlof, 1970), it is reasonable that the inconsistency attribute (DIF) discriminates solvent from insolvent HMOs. Hence, regulators should assess the quality of financial information received in order to detect underlying signals – in this case, financial distress.

7. Conclusions

HMOs assume and incur risks that are different from ordinary commercial entities. Therefore, it is expected that attributes suitable to discriminate solvent from insolvent HMOs will not be the same as those to discriminate other types of entities’ solvency.

Although the Brazilian regulatory agency of the HMO industry, ANS, does not regularly assess the quality of the financial information it receives from regulated entities, the ANS analyses all financial ratios that it considers relevant to determine an HMO’s solvency status.

As presented in prior literature, organizations that are not comfortable with their financial numbers, and suffer significant constraints from a regulatory board, have incentives to respond with misleading accounting statements. Hence, the intrinsic quality of financial reports tells more about the firms’ performance and financial position than the content of financial numbers by themselves.

This work used financial statements collected from 2,033 Brazilian Health Management Organizations and introduced a binary variable, named DIF, which is used to categorize the level of inconsistencies found in those statements. This attribute was then used, together with other traditional attributes based on ratios, to classify the HMOs either as solvent or insolvent. The results show that the use of the attribute DIF improves the classification accuracy by 5.1 percentage points, from 69.1% to 74.2%. Therefore, this study has shown that data containing inconsistencies should not be discarded, which normally occurs. On the contrary, if inconsistencies are used in a well-thought way, they can have a positive impact in data analysis, as shown here for the specific case of characterizing the solvency status of companies.

References

Altman E.I., Saunders A. Credit risk measurement: developments over the last 20 years. Journal of Banking and Finance, 21(11-12):1721-1742, 1998
Cormier, D, Magnan, M, Morard, B. The contractual and value relevance of reported earnings in a