Financial health and risk in the tourism sector in Colombia

F. Juárez

Abstract—They were applied a multivariate model of health, and a risk factors analysis within an epidemiological model to the financial statements. The sample was composed of the companies in the tourism sector in Colombia, the financial variables of cash flow, profit and loss and ownership equity were used as health indicators. Scores on these indicators were normalized and used for grouping the companies into two groups (healthy and diseased) by a hierarchical cluster. They were also estimated cutoff points in other different variables on the financial statements related to the groups obtained, by optimal binning. Finally, the odds ratio and relative risk, to determine the association between the presence or absence of disease risk and financial health of companies, were obtained. This showed the existence of different variables in the financial statements, which can divide the companies into healthy and diseased, as the risk factors.

Keywords—Financial statements, financial health, risk factors, cash flow, profit and loss, ownership equity.

I. INTRODUCTION

At times, it has been highlighted the complexity of economic processes [1], using multi-criteria models in the analysis of cash flow [2] and decision making [3], among others.

Organizations can be regarded as complex [4] and living systems [5], where health indicators acquire a complex perspective [6]. Health can be defined by global indicators summarizing many factors [7] and integrating different concepts of health [8, 9, 10] at the individual level [11]. In this way, organizations, conceived as living organisms have a number of measures of health status, among them are those related to financial activity.

The financial statements of companies can be summarized into global indicators such as the balance sheet (ownership equity, or assets minus liabilities), cash flow (i.e. cash flow of the present year) and the income statement (profit and loss). These indicators have been considered as general characteristic of financial performance of organizations. They comprise a wide range of accounts, which are aggregated into these global indicators.

Due to the fact that these three indicators emerge from different accounts in the financial statement, and also because the whole company's financial activity is included in these accounts, these three indicators are, in some way, related to each other, showing intricate relationships.

In this sense, the concept of health is composed of multiple dimensions also related one to another. Besides, health and sickness are not just opposite, but some degree of both concepts can be observed in every individual, adding a sort of complexity to health. In this way, health is a multidimensional and complex concept.

Bearing in mind the metaphor of the organization as a living entity, capable of adopting health and disease states, it is possible to use financial indicators of ownership equity, cash flow and profit and loss, as health indicators. This can be accomplished from the perspective of health and illness as a statistical process.

In previous studies, it was observed that the use of a health model and the theory of chaos provided a good predictability of the relationship among different aggregated financial indicators in the industry of crude oil and natural gas [12] and in the health sector [13]. These studies considered the indicators of cash flow, profit and loss and equity ownership as dimensions of health, and by logarithmic transformations and applying the Lorenz equations of chaos to data it was obtained a distribution which allowed for a good prediction by linear or quadratic functions.

In this studies it was adopted a continuous health perspective, a dimension where to place the different companies of the sectors. However, it is possible to understand health and disease as states that allow to classify individuals into groups according to the degree they possess those characteristics.

Moreover, within this perspective, the concept of risk assumes great importance. This concept has been used several times in various financial fields, for example, to characterize the processes of business interruption and the asymmetry generated in the market [14], to establish credit risk by semi-Markov models [15], or to characterize the growth-oriented companies [16].

In another study, qualifying companies as successful or not in the insurance sector, depended on their location above or below the geometric mean and median, considering production as an indicator and showing that a logistic regression model was appropriate to explain the classification obtained [17]. This has to do with the view of companies as healthy or not and with the risk factors associated.

It is also possible to analyze whether companies can be classified according to indicators in the financial statement, and whether the influence of internal risk factors of financial
In Colombia, the Superintendence of Societies has identified some of the reasons that lead companies to suspend payment of its obligations and subject to reorganization, declaring them as insolvent. These causes include the revaluation of the colombian currency (peso), competition and smuggling, high debt, the reduction in sales, poor organizational, administrative and financial systems, lack of working capital, problems with strategic suppliers, natural phenomena and low portfolio turnover, among others [18].

However, these causes are taken from the report issued by employers to qualify for the reorganization process, which makes the consideration of such causes as risk factors to acquire a great subjectivity and being contaminated both by the need to invoke the process and the perception of the employer at that time.

This makes it necessary to investigate other factors that constitute risks. The perspective of health and epidemiological models could help to provide a different light by adding new information.

## METHODS

The financial statements of 150 companies in the tourism sector in Colombia, for the year 2008, according to the information provided by the Superintendence of Societies, were taken as a sample. No difference was made by type of activity within the sector, company size or other characteristics.

For the overall indicator of financial health of the companies, they were used the variables cash flow of the present year, gains and loss and ownership equity. Analyses were performed using the Statistical Package for Social Sciences (SPSS).

## RESULTS

Table 1 shows descriptive data for cash flow of the present year, gains and loss and ownership equity, in the tourism sector.

This sector does not have the resources of other successful sectors of the Colombian economy, but also it does not suffer from the lack of resources from other less fortunate.

<table>
<thead>
<tr>
<th></th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ownership equity</td>
<td>-1010618</td>
<td>8987346</td>
<td>682254.54</td>
<td>1085288.230</td>
</tr>
<tr>
<td>Cash flow this year</td>
<td>0</td>
<td>3607850</td>
<td>342974.97</td>
<td>568735.391</td>
</tr>
<tr>
<td>Profit and loss</td>
<td>-1570126</td>
<td>2202603</td>
<td>90876.32</td>
<td>323978.878</td>
</tr>
</tbody>
</table>

In thousands of colombian pesos

Within the statistical model of health, it is assumed a multivariate normal distribution, where the health status is situated [10]. To accommodate to this model, the scores of cash flow of the present year, gains and loss and ownership equity, were normalized and transformed into a standardized scale, with a mean of 50 and a standard deviation of 10 (T scores). The normalization of the scores was performed using Blom's method, which gives a better fit than the direct standardization of scores. In this formulation $R$ is the range of the score and $W$ is the total number of scores. Accordingly, for the dimensions of cash flow of the present year, gains and loss and equity ownership, it was applied the following Blom's transformation [19]:

$$B = \left(\frac{R}{W - \frac{1}{4}}\right) \left(\frac{1}{\Phi^{-1}}\right)$$

The T-score transformation was performed by

$$T = z(\Phi^{-1}B) 10 + 50$$

where $\Phi^{-1}$ is the inverse cumulative normal function and $z$ is the standard deviation for each rank in the distribution.

In this multivariate normal distribution, it is necessary to set cut-off points to differentiate healthy companies of non-healthy or sick [20]. This could be done in the financial statements using the usual ratios, related to solvency, liquidity or sustainability, among others. These ratios are accompanied by some interpretation rules to decide which company is in a good financial state and, sometimes, they tend to be based in observations of the companies who have been forced to cease trading.

However, on this occasion, a statistical approach was used, which does not assume a particular cut-off point, but that point is obtained statistically. Accordingly, and in the absence of the absolute cut-off point to apply to the new standard scores, it was used a hierarchical clustering procedure, with the clustering method of Ward and the squared of the euclidean
distance. This method maximizes the difference between the groups obtained.

Euclidean distance has been used occasionally in the analysis of financial statements, as in studies of accounting harmonization among different countries [21]. It is a simple measure that is calculated by the following equation, assuming that \( x, y \) belong to a space of points or measures

\[
ED(x, y) = \sum_i (x_i - y_i)^2
\]

(3)

Using the Ward’s method companies were grouped according to the distance among them, by the following formulation [22]:

\[
W(A, B) = \frac{p_A p_B d^2(g_A - g_B)}{p_A + p_B}
\]

(4)

where \( p_A \) and \( p_B \) are the number of points in each cluster, \( g_A \) and \( g_B \) are the centers of the clusters and \( d^2 \) is the euclidean distance.

The procedure searches iteratively the pair of clusters that have the greatest similarities, these clusters are merged, the distance matrix is updated and so on, to produce a single final cluster. The researcher decides the number of clusters that considers most appropriate, according to the purpose of research.

This procedure allowed classifying companies by their financial health, according to the statistical approach to health [23], which establishes that health or illness should be determined based on statistical classification of individuals.

The classification thus obtained accounted for a distribution in two groups, as it is shown in Figure 1. In this figure, the actual values of variables, or the indicators of health, i.e. cash flow of the present year, profit and loss and ownership equities, are shown, instead of the normalized and standardized values which were only used for classification purposes.

![Figure 1. Clusters of companies according to financial indicators. Cluster 1: Sick companies. Cluster 2: Healthy companies.](image)

In the figure, we can see clearly how the variables of cash flow of the present year, profit and loss and ownership equity, allowed the classification of companies in an orderly manner, according to the simultaneous increase in these three indicators of financial health.

Based on the grouping showed in Figure 1, and in the concept of health, it can be assumed that group one is the group with symptoms of disease, due to the small scores of its financial indicators, and group two is the healthy group, due to the high scores of its financial indicators. This is a relative
classification based on the comparison between the two
groups, and not in any absolute criteria.

Considering the state of health or disease, such as in this
grouping, there is the possibility of using classical
epidemiological analysis of the influence of risk factors, on the
belonging to one group or another. Risk factors are considered
those that increase the likelihood of illness, while protective
factors are considered those that increase the chance of being
healthy.

The groups in the Figure 1 show an uneven distribution of
the companies, as in group one (sick) are located 123
companies and in group two (healthy) are only 23 companies.
This distribution and the difference between the scores of both
groups show the dominance of a small number of companies,
on the others. However, it is noted that this grouping has been
made seeking to force a classification into two groups that
maximize the differences between clusters and because of the
epidemiological model that will be apply later. With another
purpose, a classification in a larger number of clusters could
have been obtained.

The variables in the financial statements conceptualized as
risk factors were: 1) 11 Available subtotal, 2) 13 Debtors
subtotal, 3) 14 Inventories subtotal, 4) 17 Deferred
Subtotal, 5) Current total assets, 6) 13 Long-term receivable
subtotal, 7) 16 Intangible subtotal, 8) 17 Deferred subtotal, 9)
18 Other asset subtotal, 10) 19 Valuations subtotal, 11) Non-
current asset total, 12) Total assets, 13) 23 Payable account
subtotal, 14) 26 Estimated liabilities and provisions subtotal,
15) 28 Other liabilities subtotal, 16) 29 Bond and
commercial paper subtotal, 17) Current liability total, 18) 23
Payable account subtotal, 19) 26 Estimated liability and
provision subtotal, 20) 27 Deferred subtotal, 21) 28 Other
liability subtotal, 22) Total non-current liabilities, 23) Total
liabilities, 24) 31 Social capital subtotal, 25) 32 Capital
surplus subtotal, 26) Gross profit, 27) Operating income, 28)
Income before taxes and adjustments for inflation, 29) Partial
subtotal, 30) Net cash flow from operating activities, 31) Net
cash flow from investing activities, 32) Net cash flow from
financial activities, and 33) Cash last year.

To obtain the effect of these variables, considered as risk
factors, the values in them were grouped into two categories
belonging to one group or another. Risk factors are those
that increase the likelihood of illness, while protective
factors are considered those that increase the chance of being
healthy.

This statistical procedure yields an indicator of information
gain (G) associated with the cut-off point as follows [24]:

\[
Gain(A \mid T; S) = Ent(S) - E(A \mid T; S) \tag{5}
\]

Ent is the entropy of class, defined by

\[
E_{\text{int}}(S) = - \sum_{i=1}^{k} P(C_i, S) \log_2(P(C_i, S)) \tag{6}
\]

with \(k\) classes \(C_i\) a \(C_k\) y \(P(C_k, S)\) or the proportion
of observations of \(S\) within class \(C_k\).

\(E\) is the class information entropy defined by

\[
E(A, T; S) = \frac{|S_1|}{|S|} E_{\text{int}}(S_1) - \frac{|S_2|}{|S|} E_{\text{int}}(S_2) \tag{7}
\]

According to (5), optimal binning accepts the classification
induced by a certain cutoff point only if [24]

\[
Gain(A \mid T; S) > \frac{\log_2(N - 1)}{N} \cdot \frac{\Delta(A \mid T; S)}{N} \tag{8}
\]

where

\[
\Delta(A, T, S) = \log_2(3^k - 2) - [ k \ Ent(S) - k_1 \ Ent(S_1) - k_2 \ Ent(S_2)]
\]

being \(k\) the number of classes in the subset \(S\).

Besides, the coefficient of entropy helps to determine the
appropriateness of the grouping, this coefficient takes into
account that a scale variable \(A\) is grouped in categories \(I\) from
another categorical variable \(C\), with \(Ai\) belonging to the \(i^{th}\) bin
[24].

The model entropy is a measure of the predictive accuracy of
an attribute \(A\) binned on the class variable \(C\). Given a set of
instances \(S\), suppose \(A\) is discretized into \(I\) bins Given \(C\),
where, the \(i^{th}\) bin has the value \(Ai\). Letting the subset of
instances in \(S\) with the value \(Ai\), the model entropy is defined
as [24]:
The lower the coefficient of entropy the better the model fits. According to this, the grouping made out of these new variables was adequate with appropriate coefficients of entropy, ranging between 0.115 and 0.579, which are satisfactory, guaranteeing the adequacy of the model.

This statistical procedure also provides a measure of threshold for each variable, i.e. the actual value of the variable that differentiates between the scale data assigned to one category of the categorical variable and those assigned to the other category of the categorical variable, i.e. the grouping obtained by binning the scale data of each variable.

However, in this grouping, some variables were not significantly different between the two groups, so they could not be categorized. It means their values were homogeneously distributed between the two categories, so no difference was found.

These variables were: 1) 14 Inventory subtotal st, 2) 17 Deferred subtotal st, 3) 13 Long-term receivable subtotal, 4) 16 Intangible subtotal, 5) 18 Other asset subtotal, 6) 19 Valuation subtotal, 7) 26 Estimated liability and provision subtotal st, 8) 29 Bond and commercial paper subtotal, Cp, 9) 23 Payable account subtotal, 10) 26 Estimated liability and provision subtotal, 11) 27 Deferred subtotal, 12) 28 Other liabilities subtotal, 13) Total non-current liabilities, 14) 32 Capital surplus subtotal, 15) Partial subtotal, 16) Net cash flow from investing activities, and 17) Net cash flow from financial activities. These variables could not be considered risk factors.

Thus, the variables whose values were spread randomly into the two categories were not considered as significant, or risk factors and they were excluded from posterior analysis. The variables whose values yielded a significantly different distribution between the two groups were considered risk factors.

These risk factors have two categories (present or absent) depending on the threshold. If the values are under the threshold, then the risk factor is present, if they are over threshold, then the risk factor is absent. According to this, the variable can constitute itself a risk factor, but its presence or absence is defined by the position of the actual value in relation to the threshold.

Similarly, health indicators lead to the two categories shown in Figure 1, which can be considered as a group of healthy companies and a group of sick companies.

Thus, rearranging data in a 2x2 table leads to

\[
E_{\text{m}} = \sum_{j=1}^{J} P(A_j) \left( -\sum_{j=1}^{J} P(C_j|A_j) \log_2 P(C_j|A_j) \right) \]  \hspace{1cm} (9)

where

\[
P(A_i) = \frac{|S_i|}{|S|} \text{ and } P(C_j|A_i) = \frac{P(C_j,A_i)}{P(A_i)}, P(C_j,A_i) = P(C_j) \]  \hspace{1cm} (10)

According to the distribution of frequencies in the table it is possible to calculate the odds ratio and the relative risk for each group in each risk factor, according to the model of classical analytical epidemiology in health, within a cross sectional study.

Odds ratio is a measure of the strength of the association between the presence of an event and having a characteristic in one specific group, and the absence of this event and having another characteristic in another group. Event and characteristic are dichotomous, as the presence or absent of the risk factor (event) and being sick or healthy (characteristic).

The odds ratio is calculated as follows

\[
OR = \frac{ad}{bc} \]  \hspace{1cm} (11)

Another epidemiological index is relative risk, which is the ratio of the probability of having a specific characteristic in the presence of an event to the probability of having that characteristic in the absence of an event.

The relative risk for each group is

\[
\text{Sick} = \frac{(a/a+b)}{(c/c+d)} \]  \hspace{1cm} (12)

\[
\text{Healthy} = \frac{(b/b+a)}{(d/c+d)} \]

Thus, it is possible to calculate the odds ratio and the relative risk for each group in each risk factor, from an analytical epidemiology perspective.

According to this, these indicators are shown in Table 2, for each of the variables used as risk factor and the healthy and diseased groups.

Besides, the table shows the original cut-off values of the scale variables, or risk factors, which differentiates healthy from sick companies.

In the table, it is also shown the Chi square test of significance for frequencies, which results to be highly significant for all of the variables. This is not surprising, as this variables were rearrange into categories by the optimal binning statistical procedure, which provides a significance classification.

In this table, we must take into account that a relative risk equal to one indicates no association, if it is greater then it is a risk factor and if lower it becomes a protective factor. Data are presented for each group of companies, diseased and healthy.
The odds ratio is depicted for each one of the variables, meaning the disparity between the presence / absence of risk factor and the companies that are healthy and sick. As it was mentioned, it is a measure of the effect and precision of the association; the higher it is the greater effect is produced.

None of the confidence intervals for these indicators had values of one inside them, which would indicate that that indicator would be not significant.

Actually, the odd ratios for all of the risk factors are quite high, so risk factors have a real incidence in every group of companies. The relative risk is over one for all the factors for the group one, the diseased group. This means that the accounts in the table are under the cut-off value and they are risk factors for this group. On the contrary, the relative risk is under one for the group two, the healthy group. This means that the accounts in the table are over the cut-off point and they are protective factors for this group.

As we can see, the difference comes from the cut-off point established for each variable, or account. This point determines whether a company is in one or another group, in this sense if a company has its financial indicators over the cut-off points, this company will be in the second group, or healthy group, but if its financial indicators are under the cut-off points, the company will be in the first group, or diseased group.

Table 2. Odds ratios and relative risk for the association between risk factors and sick-healthy companies.

<table>
<thead>
<tr>
<th>Cut-off points for sick-healthy companies</th>
<th>Odds ratio (Sick)</th>
<th>Relative risk (Sick)</th>
<th>Relative risk (Healthy)</th>
<th>Chi squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>11 Available subtotal</td>
<td>502171</td>
<td>112.125</td>
<td>4.674</td>
<td>0.042</td>
</tr>
<tr>
<td>13 Debtors subtotal st</td>
<td>1070217</td>
<td>74.000</td>
<td>2.921</td>
<td>0.039</td>
</tr>
<tr>
<td>Total current assets</td>
<td>1800192</td>
<td>179.688</td>
<td>4.054</td>
<td>0.023</td>
</tr>
<tr>
<td>17 Deferred subtotal</td>
<td>57829</td>
<td>20.000</td>
<td>3.478</td>
<td>0.174</td>
</tr>
<tr>
<td>Total non-current assets</td>
<td>110352</td>
<td>1.482</td>
<td>26.579**</td>
<td></td>
</tr>
<tr>
<td>Total assets</td>
<td>2294396</td>
<td>264.727</td>
<td>3.334</td>
<td>0.013</td>
</tr>
<tr>
<td>23 Payable account subtotal st</td>
<td>97346</td>
<td>18.595</td>
<td>1.593</td>
<td>0.086</td>
</tr>
<tr>
<td>28 Other liabilities subtotal st</td>
<td>193705</td>
<td>22.750</td>
<td>1.694</td>
<td>0.074</td>
</tr>
<tr>
<td>Total current liabilities</td>
<td>886292</td>
<td>151.667</td>
<td>2.421</td>
<td>0.016</td>
</tr>
<tr>
<td>Total liabilities</td>
<td>1178347</td>
<td>77.941</td>
<td>3.425</td>
<td>0.031</td>
</tr>
<tr>
<td>31 Social capital subtotal</td>
<td>250000</td>
<td>23.406</td>
<td>2.190</td>
<td>0.094</td>
</tr>
<tr>
<td>Gross profit</td>
<td>1821177</td>
<td>101.333</td>
<td>3.573</td>
<td>0.025</td>
</tr>
<tr>
<td>Operating income</td>
<td>198130</td>
<td>72.833</td>
<td>3.435</td>
<td>0.047</td>
</tr>
<tr>
<td>Income before taxes and inflation adjustments</td>
<td>195868</td>
<td>7.750</td>
<td>126.397**</td>
<td></td>
</tr>
<tr>
<td>Net cash flow from operating activities</td>
<td>637879</td>
<td>85.800</td>
<td>4.475</td>
<td>0.052</td>
</tr>
<tr>
<td>Cash flow last year</td>
<td>439044</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Another interpretation of the data in the table could be that the odd ratio provides the strength of the association between the event and the characteristic of the groups and the relative risk determines the extent to which this is true. In some way, relative risk indicates the probability for a company to be in the first group, if its financial indicators are under the cut-off points (exposed to the risk factor), compare to a company which is not. This probability is lower for the second or healthy group.

IV. CONCLUSIONS

Several concepts have been introduced in this paper. These concepts come from different disciplines, like management, health or epidemiology.

Management theory has long been interested in the use of explanations from the sciences of complexity, which are far from linear models, seeking more precise explanations of facts that are difficult to model with other standard techniques.

This has been done, for example, in organizational theory [25], in studying the interdependence between firms and groups [26], the justification for the use of the complexity in management perspective [27], the organizational dynamics
or the conception of organizational culture as a complex system [29] among other applications.

Likewise, the health sciences have been concerned with integrating different explanations to provide a perspective of health and illness that involves both the dichotomy between the two states and the possibility that this dichotomy be only a part of reality. Therefore, there are much more flexible aspects and even convergence between health and disease. The links between health and disease have been described in different ways [30, 31, 32, 33, 34, 35], revealing the intricacies of them.

The fact that health and disease may also be related phenomena, depending on population characteristics, makes these concepts not to be merely static but changing.

Thus, there is a convergence between the two disciplines in the sense that they seek explanations that can account for more complex or complicated phenomena and the use of a common language, which also facilitates the rapprochement between the two.

However, the use of language of health in the science of administration must be done within limits that that science can accept. Thus, this study has features that offer some opportunities to better facilitate this approach. These features are conceptual and methodological.

Among the conceptual issues, they are the use of a statistical conception of health and disease, followed by a dichotomous conception. This approach favors using the concept of health in its more operational and pragmatic focus, without resorting to metaphors of health sciences that are difficult to accommodate in the administration science, although they may be attractive. In this sense, the metaphor has been used only as to the terms of health and disease and its operationalization in terms almost numerics.

Since health has many meanings, some of them closely related to biological components, it seems strange to use some metaphors without a careful analysis of the relationship between disciplines and without seeking terms that can be integrated into both. This has been avoided in this research.

Moreover, in this paper it has been proposed a statistical analysis procedure, which is in line with the concepts of health and disease, and is usually used in many analyses. Although previous studies had suggested using a logarithmic transformation and application of the formulas of Lorenz in the chaos theory [12] [13], it is much easier to interpret the results obtained here than those in these studies. This is due to the fact that the use of a cutoff point and risk factors around this point is intuitively understandable.

This analysis was also distinct from the manner in which the risk factors are often described in the literature is administration. They tend to be established based on a posteriori criteria, as in the case of Colombia [18] and not as dependent on the mobility criteria of the sector or sectors, as it has been done in this research.

In this sense it is noteworthy that although the techniques used in this study are commonly used and have been developed for some time ago, however, it is the combination of the techniques within a particular conceptual framework which makes them different. That is, the description of risk factors is performed within a mobile concept of health depending on the population or sector and the time at which analysis is undertaken.

Such a combination provides changing criteria and groupings that depend on the dynamics of the sector or the economy, and the results are different from one moment to another. Thus, the organization as a living entity is subject to a constantly changing definition of financial health and, as a result of the statistical definition of it, there will always be healthy and sick companies.

Even more, the financial condition could not be considered as a state of disease if a sufficient number of companies are under the same conditions, as this would result in a grouping where they could be considered healthy those companies with a particular state.

All this acquires a certain complexity and is relevant to a detailed conceptual analysis. In addition, the analysis techniques used here are much more likely to be refined through this combination of analysis, based on health and epidemiological models, to provide a more dense explanation.

Moreover, it is also necessary to include health, epidemiological and administrative models that explain why some variables in the financial statements have become risk factors, which has not been done in this paper.

Finally, it can be concluded that the application of health and epidemiological models can be of interest in analyzing financial statements. However, it is needed to go deeper into their application, incorporating the use of more complex concepts and calculations, and the usual ratios of financial analysis.  

REFERENCES
