ARTICLE IN PRESS

Advances in Accounting xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Advances in Accounting

journal homepage: www.elsevier.com/locate/adiac

Financial performance and distress profiles. From classification according to financial ratios to compositional classification

Salvador Linares-Mustarós[a], Germà Coenders[b,*], Marina Vives-Mestres[c]

[a] Department of Business, University of Girona, Campus Montilivi, Faculty building of Economics and Business, C. Universitat 10, 17003 Girona, Spain
[b] Department of Economics, University of Girona, Campus Montilivi, Faculty Building of Economics and Business, C. Universitat 10, 17003 Girona, Spain
[c] Department of Computer Science, Applied Mathematics and Statistics, University of Girona, Campus Montilivi, Building P4. C. Maria Aurèlia Capmany, 61, 17003 Girona, Spain

ARTICLE INFO

Keywords:
Compositional data analysis (CoDa)
Financial ratio
Cluster analysis
Ilr coordinate
Aitchison's distance

JEL classification:
C49
G30
M41

ABSTRACT

Financial ratios are often used in cluster analysis to classify firms according to the similarity of their financial structures. Besides the dependence of distances on ratio choice, ratios themselves have a number of serious problems when subject to a cluster analysis such as skewed distributions, outliers, and redundancy. Some solutions to overcome those drawbacks have been proposed in the literature, but have proven problematic. In this work we put forward an alternative financial statement analysis method for classifying firms which aims at solving the above mentioned shortcomings and draws from compositional data analysis. The method is based on the use of existent clustering methods with standard software on transformed data by means of the so-called isometric logarithms of ratios. The method saves analysis steps (outlier treatment and data reduction) while defining distances among firms in a meaningful way which does not depend on the particular ratios selected. We show examples of application to two different industries and compare the results with those obtained from standard ratios.

1. Introduction

Financial ratios, i.e., ratios comparing the magnitudes of accounts in financial statements, constitute a case of researchers' and professionals' interest in relative rather than absolute account magnitudes. From the classical work on bankruptcy prediction by Altman (1968), the use of financial ratios has spread along and across many research lines (Willer do Prado et al., 2016), such as stock market returns (e.g., Dimitropoulos, Asteriou, & Koumanakos, 2010), firm survival analysis (e.g., Kalak & Hudson, 2016), credit scoring (e.g., Amat, Manini, & Antón Renart, 2017), assessing the impact of International financial reporting standards (e.g., Lueg, Punda, & Burkert, 2014), predicting donations to charitable organizations (e.g., Trussel & Parsons, 2007), accounting restatements (e.g., Jiang, Habib, & Zhou, 2015), and earnings manipulation (e.g., Campa, 2015). This article focuses on another frequent use of financial ratios: to classify firms according to similarity of the structure of their financial statements, searching for different profiles of financial structure, performance or distress. Since the seminal works of Cowen and Hoffer (1982), and Gupta and Huefner (1972), through the relevant contributions by Dahlstedt, Salmi, Luoma, and Laakkonen (1994); Ganesalingam and Kumar (2001); Mar Molinero, Apellaniz Gomez, and Serrano Cinca (1996); Serrano Cinca (1998); and Voulgaris, Doumpos, and Zopounidis (2000), the interest in clustering firms according to their financial ratios remains current (Feranecová & Krigovská, 2016; Lukason & Laitinen, 2016; Luptak, Boda, & Szucs, 2016; Martín-Oliver, Ruano, & Salas-Fumás, 2017; Momeni, Mohseni, & Soofi, 2015; Santis, Albuquerque, & Lizarelli, 2016; Sharma, Shebalkov, & Yukhanaev, 2016; Yoshino & Taghizadeh-Hesary, 2015; Yoshino, Taghizadeh-Hesary, Charoensivakorn, & Niraula, 2016).

Despite the popularity of financial ratios, the financial and statistical literature has long reported a number of serious practical drawbacks of their use. The first of them has to do with the fact that most ratios are distributed between zero and infinity and thus make fully symmetric distributions impossible to achieve. Ratios also tend to have asymmetric distributions because decreases in the denominator produce larger changes in the ratio value than increases do (Frecka & Hopwood, 1983). Both phenomena tend to produce distributions with positive skewness and preclude using symmetric probability distributions such as the normal (e.g., Deakin, 1976; Ezzamel & Mar-Molinero, 1990; Kane, Richardson, & Meade, 1998; Martikainen, Perttunen, Yli-Olli, & Gunasekaran, 1995; Mcleay & Omar, 2000; So, 1987). Asymmetry is also connected to the commonly reported outliers (e.g., Cowen & Hoffer, 1982; Ezzamel & Mar-Molinero,

* Corresponding author.
E-mail addresses: salvador.linares@udg.edu (S. Linares-Mustarós), germa.coenders@udg.edu (G. Coenders), marina.vives@udg.edu (M. Vives-Mestres).

http://dx.doi.org/10.1016/j.adiac.2017.10.003
Received 13 September 2017; Received in revised form 16 October 2017; Accepted 25 October 2017

1990; Lev & Sunder, 1979; So, 1987; Watson, 1990). It can even be the case that outliers are the main or only source of positive asymmetry in the distributions (Frecka & Hopwood, 1983). These outliers do not always reflect atypical management practices but can also result from a small value of the denominator of the ratio (e.g., Ezzamel & Mar-Molinero, 1990; Kane et al., 1998). In the particular case of cluster analysis, asymmetric distributions lead to some clusters being very small (e.g., Feranecová & Krigovská, 2016; Santis et al., 2016; Sharma et al., 2016; Yoshino et al., 2016; Yoshino & Taghizadeh-Hesary, 2015). It is also well known that the presence of outliers distorts the results of many clustering algorithms, and oftentimes it even leads to one-member clusters (e.g., Feranecová & Krigovská, 2016; Sharma et al., 2016; Yap, Mohamed, & Chong, 2014).

The second major drawback has to do with redundancy of the over 100 ratios currently in use (Chen & Shimerda, 1981; Pindado & Rodrigues, 2004; Pohlman & Hollinger, 1981). Oftentimes, redundancy occurs to such an extent that "there is no absolute test for the importance of variables" (Barnes, 1987, 455) and "to identify those ratios which contain complete information about a firm while minimising duplication cannot be achieved purely by logic" (Barnes, 1987, 456). In extreme cases there is an exact dependency between ratios. For instance, the inverse of the liability to asset ratio is the equity to debt ratio plus one. In cluster analysis such redundancy has often led to groups not capturing proper distinct profiles. Solutions with overall ordered groups labelled as "healthy, in between, less healthy", "highly distressed, mildly distressed, not distressed", "dynamic, medium, weak", or "good performers, average performers, poor performers" abound (e.g., Ganesalingam & Kumar, 2001; Momeni et al., 2015; Voulgaris et al., 2000; Yap et al., 2014; Yoshino et al., 2016; Yoshino & Taghizadeh-Hesary, 2015). In cluster analysis, redundancy has one further consequence: it increases distances among firms along the added redundant information, which is tantamount to inadvertently giving this redundant information greater weight in the results (e.g., Aldenderfer & Blashfield, 1984).

The third major drawback has to do with arbitrariness of Euclidean distance among firms. On the one hand, a different set of ratios leads to different distances among firms, even if ratios are computed from exactly the same set of financial accounts. On the other hand, Euclidean distance is not an appropriate dissimilarity measure for ratios. Even placement of accounts in the numerator or in the denominator of the same ratio matters to Euclidean distance. This is because increases in the numerator and in the denominator are not treated in the same way (Frecka & Hopwood, 1983). Let us consider the simplest possible case in which only two financial accounts x_1 and x_2 are of interest. Only two ratios are possible: $r_1 = x_1/x_2$ and $r_2 = x_2/x_1$. Let us consider three firms A, B, and C, such that $x_{1A} = 1$, $x_{2A} = 1$, $x_{1B} = 1$, $x_{2B} = 2$, $x_{1C} = 2$, $x_{2C} = 1$. The ratio values are $r_{1A} = r_{2A} = 1$, $r_{1B} = 0.5$, $r_{2B} = 2$, $r_{1C} = 2$, $r_{2C} = 0.5$. Intuitively, the ratios r_1 and r_2 should contain the same information about firms. However, Euclidean distances computed from r_1 are d(A,B) = 0.5, d(A,C) = 1, and d(B,C) = 1.5, while Euclidean distances computed from r_2 are d(A,B) = 1, d(A,C) = 0.5, and d(B,C) = 1.5. In other words, when using r_1 firms A and B would tend to cluster together and when using r_2 firms A and C would tend to cluster together. Unclear distances which depend on arbitrary decisions and even on a permutation of numerator and denominator can only threaten the results of cluster analysis (Martín, 1998).

As regards the problem related to asymmetry and outliers, some form of transformation and/or outlier trimming has often been applied. These include transformations such as Box-Cox (e.g., Ezzamel & Mar-Molinero, 1990; Mcleay & Omar, 2000; Watson, 1990), logs (e.g., Cowen & Hoffer, 1982; Deakin, 1976; Sudarsanam & Taffler, 1995), ranks (e.g., Kane et al., 1998; Lueg et al., 2014), square roots (e.g., Deakin, 1976; Frecka & Hopwood, 1983; Martikainen et al., 1995), weight of evidence (e.g., Nikolic, Zarkic-Joksimovic, Stojanovski, & Joksimovic, 2013); outlier trimming (e.g., Ezzamel & Mar-Molinero, 1990; Frecka & Hopwood, 1983; Lev & Sunder, 1979; Martikainen et al., 1995; So, 1987; Watson, 1990); and outlier winsorization (e.g., Lev & Sunder, 1979).

Both transformation and outlier treatment have proved problematic. Not only is there uncertainty about which transformation to apply or which outliers to remove. There is also uncertainty regarding whether one should first remove outliers and then transform to account for the remaining non-normality or first transform and then remove the remaining outliers (e.g., Ezzamel & Mar-Molinero, 1990). The log transformation is especially appealing, given its wide understanding and ease of interpretation as relative change in the economic and financial fields. It is also theoretically justified when the numerator and the denominator follow a log-normal distribution. Empirically it is also often reported to yield acceptable results (Sudarsanam & Taffler, 1995). However, as shown above, there is no consensus on the transformation issue, and in some cases more than one transformation has been shown to yield approximately normal ratios (Buijink & Jegers, 1986).

As regards the redundancy problem, many clustering studies use data reduction methods prior to the analysis, either to compute a few aggregated functions of ratios or to select a few relevant and distinct ratios. These strategies include principal component analysis (e.g., Cowen & Hoffer, 1982; Dimitropoulos et al., 2010; Martín-Oliver et al., 2017; Sharma et al., 2016; Yoshino et al., 2016; Yoshino & Taghizadeh-Hesary, 2015), grey relation analysis (e.g., Ho & Wu, 2006), factor analysis (e.g., Feranecová & Krigovská, 2016; Lukason & Laitinen, 2016; Yap et al., 2014), self-organising feature maps (e.g., Serrano Cinca, 1998), multidimensional scaling (e.g., Mar Molinero et al., 1996), or cluster analysis on the transposed data matrix, to define groups of ratios instead of groups of firms (e.g., Nikolic et al., 2013; Serrano Cinca, 1998). While this is generally sound practice, it adds an extra step to the analysis, and it is often not clear which data reduction method should be preferred for a particular problem.

To the best of our knowledge, the distance issue has not been solved in the financial literature, but it has been solved in other scientific fields, from which we draw below.

The aim of this article is to put forward an alternative financial statement analysis method for classifying firms from the structure of their financial statements, which aims at solving the above mentioned shortcomings and draws from the compositional data analysis (CoDa) literature. CoDa is the standard methodological toolbox to analyse the relative importance of magnitudes in fields such as biology, chemistry and geology. A key feature of CoDa is a particular type of log transformation of ratios, which tends to lead to symmetric distributions with few or no outliers, and to less redundancy, thus making data reduction less necessary. This transformation also ensures that the distances among clustered cases are meaningful and that they only depend on the set of financial accounts which is considered for the analysis and not on ratio choice. Once this transformation has been carried out, standard clustering methods and software can be used, which is an attractive possibility for applied researchers.

The article is organized as follows. Section 2 reviews the basics of CoDa. Section 3 deals with the proposal to use an alternative financial statement analysis method based on CoDa. Section 4 presents two numerical real-data examples of cluster analysis in high tech and low tech manufacturing industries. Results are compared to those obtained when using standard financial ratios. Section 5 summarizes the main results and makes suggestions for further research.

2. Compositional data analysis

2.1. Compositional data

Compositional Data are positive vector variables carrying information about the relative size of their D components to one another (Aitchison, 1986; Barceló-Vidal & Martín-Fernández, 2016):