J. of Mult.-Valued Logic & Soft Computing, Vol. 47, pp. 491–507
Reprints available directly from the publisher
Photocopying permitted by license only

Beyond Traditional Ratios: Predicting Financial Distress through Compositional Data and Fuzzy Clustering in Agricultural Retail

XAVIER MOLAS-COLOMER[1,*], JOAN CARLES FERRER-COMALAT[1], MARC CARRERAS-PIJUAN[1,2] AND SALVADOR LINARES-MUSTARÓS[1]

[1]*Department of Business Studies, Faculty of Economics and Management - Campus Montilivi, University of Girona, C/ Universitat de Girona, 10 - 17003, Girona, Spain*
[2]*Montilivi Campus, University of Girona, C/ Universitat de Girona, 10 - 17003, Girona, Spain, his affiliation as a Serra Húnter Fellow, Serra Húnter Programme, Barcelona, Spain*

Accepted: February 15, 2026.

Bankruptcy prediction plays a crucial role in financial risk assessment, but traditional methods based on classic financial ratios face important limitations due to asymmetry, outliers, and redundancy. This study explores whether combining Compositional Data Analysis (CoDa) with fuzzy cluster analysis provides a more reliable approach to bankruptcy prediction. Using financial data from agricultural retail firms obtained from the SABI data set with last available financial year 2024, or at least 2023, financial statements are transformed into centered log-ratios and analyzed through fuzzy clustering to identify distinct financial profiles. The resulting cluster membership values, treated as compositional data, are transformed into additive log-ratios and used as predictors in a logit regression model to estimate bankruptcy probabilities. The results indicate that the compositional approach can improve prediction accuracy by adequately addressing the relative nature of financial ratios while reducing the distorting effects of asymmetry and outliers, improving conventional methods.

Keywords: Compositional data analysis (CoDa), financial ratio, cluster analysis, fuzzy cluster, bankruptcy prediction

* Corresponding author email: xavier.molas@udg.edu

1 INTRODUCTION

The agricultural retail distribution sector in Spain is a fundamental pillar of the country's agricultural economy, as it guarantees the supply of essential goods for primary production, such as seeds, fertilizers, animal feed, and farming equipment. This sector plays a key role in the agri-food value chain, facilitating the transfer of productive inputs to farmers and contributing to food security and the sustainability of the production system.

The economic resilience of the sector is reinforced by various structural factors. On the one hand, strong domestic demand ensures a relatively stable consumption base, linked to the strategic importance of agriculture in the national context. On the other hand, European Union subsidies, mainly through the Common Agricultural Policy (CAP), provide financial support that helps mitigate the volatility of agricultural incomes and, indirectly, stabilize the distribution market. Likewise, the growing adoption of sustainable and ecological agricultural practices generates new commercial opportunities and favors the differentiation of products with greater added value [1], [2].

Despite these positive elements, the sector faces significant challenges. The volatility of agricultural input prices, often influenced by global factors, increases economic uncertainty. Added to this is the rise in production costs, resulting from higher energy, raw material, and logistics costs. Finally, competitive pressure from large international operators limits the margins of local companies and demands greater operational efficiency.

In this context, business adaptation strategies become essential. Diversifying the product offering reduces dependence on specific market segments. Investment in digital platforms improves commercial efficiency, optimizes supply chain management, and facilitates access to new sales channels. In addition, the formation of cooperatives and business alliances helps to strengthen bargaining power, generate economies of scale, and increase the sector's capacity for innovation. Taken together, these strategies are crucial to ensuring the long-term competitiveness and sustainability of agricultural retail distribution in Spain.

Bankruptcy prediction plays a crucial role in financial risk assessment by determining the likelihood of a company's insolvency. Traditional methods of financial stress monitoring typically rely on classic financial ratios for comparisons within industries in the same sector [3], [4]. The goal is to provide a reasonable probability that a company is likely to go bankrupt in the short term. In order to be able to group and to categorize firms according to the nature of their financial statements, observing different profiles of their financial structure, their performance or the problems that their balance sheet may indicate, cluster analysis using financial ratios has been increasing its popularity among the researchers [5], [6], [7], [8] and [9] .

Although traditional ratios can provide useful information in the individual analysis of companies, their use in sectorial and statistical studies has serious methodological limitations, which call into question their reliability; the inherent mathematical structure of these ratios, when treated as variables in such analyses, presents practical challenges and limitations [10], [11] and [12].

These issues have already prompted several studies on the topic. Among them, we wish to highlight the asymmetric distributions of results seen in studies using ratios. This is particularly important because these ratios typically range from zero to infinity, and a reduction in the denominator leads to a more pronounced change in the ratio than an increase in the numerator [9],[13]. Such characteristics make it inappropriate to apply symmetric probability distributions, like the normal distribution [14], [15], [16], [17] and [18]. The skewness can also give rise to misleading outliers [19], [20], [21], [22] and also [12] and [15], and these outliers may be the primary, or even the sole, cause of the positive asymmetry in the data. However, in practice, these results don't always reflect fundamentally different conditions. In the context of cluster analysis, such asymmetric distributions can lead to very small clusters. Furthermore, the presence of outliers is well-known to distort the results of clustering algorithms, sometimes producing single-member clusters [23], [24] and [25]. A second problem is related to statistical replication or redundancy of ratios, which arises more frequently in studies with a very large number of data [26], [27].

The Compositional Data (CoDa) methodology approach addresses many of the previously discussed challenges and limitations. By means of suitable logarithmic transformations, CoDa leads to symmetric distributions with few or no outliers, with less redundancy of information so that no data reduction is required, and such transformation ensures that the distances among clustered points are significant and they only depend on the set of financial accounts considered for analysis and not on the ratio chosen to analyze them, minimizing the problem of permutation of numerator and denominator in a ratio construction.

This study explores whether Compositional Data analysis, whose validity of its results has already been extensively tested in many different fields [28-34] and also [9], and fuzzy cluster analysis [35], [36] and [37] can provide an innovative bankruptcy prediction technique. Specifically, we examine the applicability of these methodologies to the agricultural retail sector, utilizing a representative dataset of financial statements from the Iberian Balance Sheet Analysis System (SABI).

The primary aim of this study is to assess whether combining CoDa and fuzzy clustering offers a more precise and insightful method for predicting bankruptcy. Unlike conventional financial ratios, compositional data provides a way to account for the relative importance of different financial components [8]. This method is particularly suited for complex datasets, where the relationships between variables are not independent but instead exhibit compositional characteristics. The approach begins with the transformation of some of their financial statements into clr log-ratios (centered log-ratio transformations), proved to be suitable as input variables for a fuzzy cluster analysis [11], which assigns companies to fuzzy clusters based on the similarity of their financial characteristics. The resulting fuzzy cluster membership values assigned to each firm, which inherently form compositional data because their sum is always one, are then converted into alr log-ratios (additive log-ratio transformations). Finally, a logit linear regression model is constructed using these alr log-ratios as variables to predict bankruptcy probabilities [38], [39] and [8].

This article is structured as follows: following the introduction, Section 2 outlines the data research and the methodology, detailing the CoDa approach to clustering financial ratios and the suitability of a fuzzy classification. Section 3 presents the results of the cluster analysis and the logit regression model applied. Section 4 presents the conclusions and discussion.

2 METHODS

We define a company as being bankrupt if fulfilling at least one of the following criteria. From an accounting perspective when liabilities exceed assets, and from a legal perspective when the firm is under *creditors' meeting* ("concurso de acreedores" in Spain). To predict if a company is bankrupt in the last available year, the financial ratio values of the year immediately before are used.

2.1 Data

The financial statements used in this study were obtained from the Sistema de Análisis de Balances Ibéricos (SABI, accessible at https://sabi.bvdinfo.com) database, developed by INFORMA D&B in collaboration with Bureau Van Dijk. The focus was placed on firms classified under the primary code of the NACE 2009: 01XX - Agriculture, livestock farming, hunting, and related services. Data from this sector with last available financial year 2024, or at least 2023, were selected. The data was obtained on October 13, 2025.

In addition to sectorial classification, firms were required to have a minimum of ten employees and a maximum of fifty to ensure the inclusion of firms in the retail sector with sufficient operational activity. Only companies registered within Spain were considered.

The selected D = 7 accounting figures were the ones needed to compute the ratios in Altman's Z score: non-current assets, current assets, retained earnings, non-current liabilities, current liabilities, operating income and operating expenses.

Filtering process were as follows: No initial outliers were be removed, only firms having either zero total assets, zero operating income and zero operating expenses during the last available year or the year before the last to be considered inactive in this period. Firms with other negative retained earnings values are set to zero. Firms with negative operating expenses were also removed to be considered inactive. The remaining zero values were handled according to the procedures exposed in Section 2.5.

The final dataset sample size was 3151 firms, of which 109 were bankrupt according to the criteria established above.

2.2 Financial statements as a composition

In the CoDa methodology, a D-part composition is defined as a set of D strictly positive numbers, called parts, where the relative magnitude of parts to one another is of interest [28], [29], [31], [32], [33], [34] and [8]. In scientific fields such as geology or chemistry, compositions often sum to a constant value, which can be defined as one, but in financial analysis, this is not required:

$$\boldsymbol{x} = \left(x_1, x_2, \ldots, x_D\right) \; with \; x_j > 0, \; j = 1, 2, \ldots, D \tag{1}$$

Two fundamental guidelines should be followed when applying CoDa to financial data. First, negative values should be excluded, as they may produce misleading interpretations and discontinuities in financial ratios (for instance, a positive return on equity arising when both profit and net worth are negative). Second, overlapping components must be avoided. As an example, including both total assets and non-current assets is inappropriate because non-current assets are a subset of total assets. One should therefore use either the complete aggregate (such as total assets) or its disaggregated components (such as current and non-current assets), but not both simultaneously.

In this study, the parts represented by the x_j variables are the following D = 7 positive and non-overlapping financial statement account categories exposed in Table 1:

We have added another non-financial binary-encoded variable y_1 in order to be able to numerically declare whether a firm was inactive the following year (1) or not (0). In the first case, and looking at the financial data for that year and immediately thereafter, we will consider the firm to be in bankruptcy.

Data from last avaliable year (2024 or 2023)		Data from 2025
x_1 = Non-current assets	x_4 = Non-current liabilities	y_1 = Bankruptcy (1:yes, 0:no)
x_2 = Current assets	x_5 = Current liabilities	
x_3 = Retained earnings	x_6 = Opertaing income	
	x_7 = Operating expense	

TABLE 1
Selected financial statement categories and bankruptcy variable

Bankruptcy	Non_current_ assets	Current_assets	Retained_ earnings	Non_current_ liabilities	Current_ liabilities	Operating_ income	Operating_ expenses
0	698,4418	379,7702	153,9892	222,7509	465,5083	1194,407	1041,755
0	607,2081	124,8957	29,4	184,2879	405,6633	798,9759	893,5925
0	30864,59	936,2455	10770,05	4,50034	1625,796	510,4373	347,8083
0	376,4644	583,7338	406,0217	95,72548	322,6545	1858,529	1648,205
0	300,849	413,824	34,8894	114,2472	459,0002	443,313	422,763
0	12259,79	8753,413	7299,163	4817,113	3898,845	26122,88	21751,98
0	14210,78	5304,057	9241,84	6126,183	1687,062	2070,231	938,4014
0	6276,137	1748,362	6548,764	67,956	325,7265	2903,274	1713,843
0	72,96414	6089,919	3913,176	0	2141,388	16761,65	16620,52
0	3163,18	1076,502	1917,957	508,1447	1101,012	3891,57	2944,604

TABLE 2
Values obtained for the first ten firms selected of the above explained financial statement account and bankruptcy variable

Table 2 shows some of the values obtained for this purpose. These seven declared parts are the most commonly used to date in the analysis of financial statements using the CoDa methodology [2], [5], [7], [22] and [39], and they fulfill the two main rules outlined above. Also, they constitute the basis for a wide array of common financial and management ratios frequently used in accounting. See, for example, *Return on Equity* (ROE), defined as operating profit over net equity:

$$\mathrm{ROE} = \frac{x_6 - x_7}{x_1 + x_2 - x_4 - x_5} \tag{2}$$

2.3 Centered log-ratios

For the application of the CoDa framework to clustering tasks, financial data must be projected into a Euclidean space by means of log-ratio transformations, which constitute a standard approach in CoDa analysis [29]. Among these, the centered log-ratio (CLR) transformation preserves the relative inter-point distances, thereby justifying the application of classical distance-based multivariate statistical methods, including cluster analysis, on the transformed data. The CLR transformation for each accounting figure in Equation 1 is defined as:

$$clr_j = log\left(\frac{x_j}{\sqrt[D]{x_1 x_2 \ldots x_D}}\right) \quad with \quad j = 1,2,\ldots,D \tag{3}$$

In financial statement analysis, each centered log-ratio involves comparing each accounting figure x_j (in the numerator) to the geometric mean of all figures for each individual firm (in the denominator). The CLRs don't have any financial interpretation in themselves, but they avoid having to manually choose the log-ratios to be used (as in pairwise log-ratios) by ensuring that the *D* CLRs also contain all the information on the relative importance of the *D* accounting values. Any log-ratio that may be of interest to the object of study is a function of these centered log-ratios. Furthermore, this transformation solves the challenges of non-linearity, asymmetry, and spurious outliers encountered in classic financial ratios [5], [6], [7], [10] and [11].

2.4 Fuzzy clustering methodology

Clustering is the multivariate statistical process of organizing a set of observations into subsets in such a way that intra-cluster similarity is maximized while inter-cluster similarity is minimized. It constitutes a fundamental methodology in statistical data analysis and is extensively applied across a wide range of scientific disciplines. Fuzzy clustering, also referred to as Soft clustering, encompasses a class of approaches in which observations are characterized by graded membership values across multiple clusters, rather than being assigned to a single cluster exclusively. The identification of clusters relies on similarity measures, which may be defined in terms of connectivity, intensity, or distance metrics [9], [11].

Fuzzy clustering algorithms were developed to address limitations inherent in exclusive clustering approaches (commonly referred to as Hard clustering), which assume that each element can be assigned unambiguously to a single cluster and exhibits no similarity to elements belonging to other clusters. This limitation was overcome through the introduction of fuzzy logic [40], which allows the degree of similarity or membership of each element to each cluster to be quantified. Such similarity is represented by a membership function that assigns values in the interval [0,1], where values closer to one denote a higher degree of similarity and values closer to zero indicate weaker resemblance. These membership values are normalized within the interval [0,1]; however, they do not necessarily correspond to a probability measure, and therefore the memberships associated with a given observation are not required to sum to one. Consequently, the fuzzy clustering problem reduces to determining an optimal membership structure that appropriately characterizes the set of data points. This fits perfectly with cluster analysis using financial statements, since the greater membership degree to a cluster of a given firm within a sector does not necessarily imply that it has no similarities with nearby firms in other clusters.

CoDa cluster analysis is reduced to performing a standard cluster analysis using the D centered log-ratios calculated using the Equation (3) as input variables. Therefore, traditional Euclidean distances correspond to Aitchison distances, which are the standard in CoDa. Consequently, the distance between two firms m and l is calculated using a Euclidean-like distance formula:

$$d_{ml} = \sqrt{\left(clr_{1m} - clr_{1l}\right)^2 + \left(clr_{2m} - clr_{2l}\right)^2 + \cdots + \left(clr_{Dm} - clr_{Dl}\right)^2} \tag{4}$$

This transformation addresses the most common shortcomings of standard financial ratios in cluster analysis, which often lead to the formation of extremely large or extremely small clusters, as well as clusters consisting of only one or two outlying observations [9], [11], [23] and [25].

One of the most widely used fuzzy-clustering algorithms in financial statement analysis is the Fuzzy K-means clustering algorithm [41], [42]. In this study, this algorithm, as implemented in R software, is applied.

2.5 Zero replacement

A common limitation working with both CoDa and classical financial ratios is that the accounting parts of interest cannot contain zero values in order to perform calculations with them: logarithms of these values cannot be operated and ratios involving zero values in the denominator are undefined. Also, a zero accounting value is not relative to anything and, therefore, the ratio with this value would not be a valid operation according to statistical measurement scale theory.

However, CoDa methodology offers advanced tools for zero imputation (also known as *zero replacement*) before log-ratio computations and under the most common statistical assumptions. This provides CoDa methodology with a clear advantage, allowing financial statements to be examined even when some relevant accounting figures are zero. Essentially, zero values are substituted with small, meaningful quantities that satisfy specific statistical properties.

The most widely applied imputation approach in the analysis of compositional financial statements is the Expectation–Maximization (EM) method for log-ratios, developed by Palarea-Albaladejo & Martín-Fernández, (2008). This approach is analogous to the standard EM algorithm used for imputing missing data, as the imputed values are estimated from the available information using a statistical model. Missing data are simply values that are absent or unknown for a given variable and firm. In the case of zero replacement in compositional data, an additional constraint is imposed: the imputed values must be “small”. Specifically, they are required to be lower than the smallest observed nonzero value of each component, or below any other user-defined threshold known as the *detection limit*.

For this study, the percentage of zeros obtained were 1.24% for non-current assets, 15.90% for non-current liabilities, 0.10% for current liabilities and 8.03% for retained earnings; all of them lower than 20%, which were acceptable values for this purpose (below the 5th percentile of non-zero values) and were imputed with the method just described. The other accounting figures had no zero values.

The zero replacement and clr transformations were carried out with CoDaPack2.03.10 (Comas-Cufí & Thió-Henestrosa, 2011). CoDaPack is freely available at https://ima.udg.edu/codapack/. An introduction to the use of CoDaPack in financial statement analysis can be checked in [8]. Further fuzzy cluster analysis, ALR transformation of cluster membership values and logit lineal regression model were carried out with R software.

2.6 Logit regression model based on fuzzy cluster memberships

The fuzzy cluster analysis yields, for each firm i, a vector of membership values expressed as follows:

$$\boldsymbol{u_i} = \left(u_{i1}, u_{i2}, \ldots, u_{ik}\right) \; with \; 0 < u_{ij} < 1, \; j = 1,2,\ldots,k \;\; and \; \sum_{j=1}^{k} u_{ij} = 1 \tag{5}$$

where u_{ij} denotes the degree of affiliation of firm i with cluster j, and k is the number of clusters. These membership values are bounded between zero and one and, by construction, sum to unity for each firm. Consequently, they belong to the simplex and must be treated as compositional data. As a result, they cannot be directly included as explanatory variables in a regression model without violating fundamental statistical assumptions, such as independence and absence of perfect multicollinearity.

To ensure a statistically coherent regression framework, the fuzzy cluster membership values must be transformed using an additive log-ratio (ALR) transformation prior to model estimation [28]. One of the clusters is selected as a reference category, and the remaining membership values are expressed as log-ratios relative to this reference. So, for each firm i, the ALR log-ratios are computed follows:

$$alr_{ij} = \log_2\left(\frac{u_{ij}}{u_{ik}}\right), \; with \;\; j < k, \; j = 1,2,\ldots,k-1 \tag{6}$$

yielding $k-1$ real-valued and unconstrained variables. The use of a base-2 logarithm preserves the relative information contained in the original memberships while facilitating interpretation. This transformation maps the compositional membership vectors from the simplex to an unconstrained Euclidean space, thereby ensuring compatibility with standard regression techniques.

The ALR-transformed membership values can then be used as explanatory variables in a binary logistic regression model, with bankruptcy status in the subsequent year defined as the dependent variable. The logit specification models the log-odds of bankruptcy as a linear function of the relative degrees of cluster affiliation. This approach allows the estimated coefficients to be interpreted in terms of changes in bankruptcy risk associated with a firm's relative proximity to each financial profile, compared to the reference cluster. Formally, the model is specified as follows:

$$\text{logit}\left(y_1\right) = \beta_0 + \sum_{j=1}^{k-1} \beta_j alr_{ij} \tag{7}$$

where $y_1 = 1$ denotes bankruptcy for firm *i*, $y_1 = 0$ otherwise, β_0 is the intercept, and β_j are the regression coefficients associated with the ALR coordinates.

From a methodological standpoint, this approach aligns with the principles of compositional data analysis and with the recommendations given in [38], who emphasize that regression models with compositional predictors should be formulated exclusively in log-ratio coordinates to ensure valid inference and meaningful interpretation. By integrating fuzzy clustering with ALR-transformed membership values in a logit model, the proposed methodology provides a coherent link between classification of financial profiles and probabilistic bankruptcy prediction.

In the logistic regression framework, the estimated coefficients associated with the ALR-transformed cluster membership values represent the effect of changes in a firm's relative affiliation to a given cluster, with respect to the reference cluster, on the log-odds of bankruptcy. A positive coefficient indicates that an increase in the relative membership to that cluster is associated with a higher probability of bankruptcy, while a negative coefficient implies a lower bankruptcy risk compared to the reference profile. The exponentiation the coefficients yields odds ratios, which quantify the multiplicative change in bankruptcy odds resulting from a unit increase in the corresponding ALR coordinate. Importantly, these effects are inherently relative, reflecting shifts in cluster affiliation rather than absolute levels of membership, in accordance with the principles of compositional data analysis.

3 RESULTS

3.1 Fuzzy cluster analysis

In most cases, the appropriate number of clusters into which a given business sector should be partitioned is not known a priori. Consequently, several statistical validity indices are commonly employed to determine the optimal

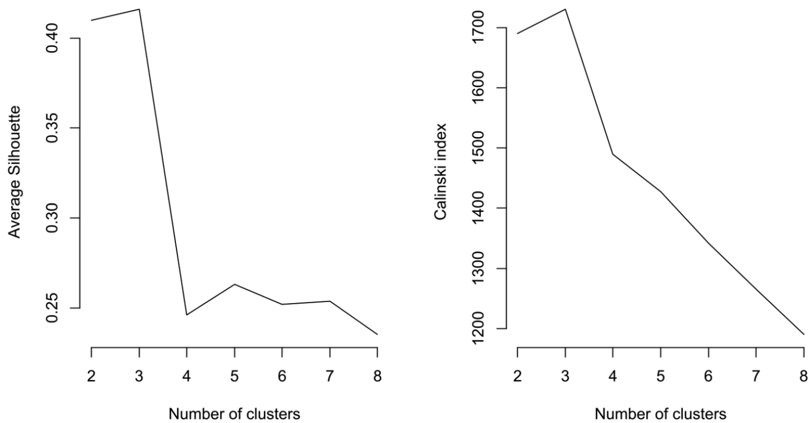

FIGURE 1
Evaluation of clustering solutions using the average silhouette width and the Calinski-Harabasz indices

number of clusters after performing classifications with a reasonable number of clusters, for example from two to eight. CoDaPack has available the *average silhouette width* [43] and the *Calinski-Harabasz* [44] indices. The larger value of both indices determines de best value of k (number of clusters).

As shown in Figure 1, in our case a three-cluster solution maximizes both criteria, so we take this $k = 3$ as the most appropriate grouping for this analysis. Also, for visualization purposes, a CoDa Biplot is generated using the 7 centered log-ratios calculated using the Equation (3),

This is only an overall perspective of how the crisp cluster classification would look; to handle fuzzy classification, we exported the CLR dataset to R to perform the Fuzzy K-means clustering algorithm in order to obtain the fuzzy cluster membership values for each firm (Table 3).

These values, inherently being compositional data because their sum is always one (sum of values in a row), are then converted into ALR (additive log-ratios).

3.2 Logit regression model

Finally, a bankruptcy for the following year logit linear regression model is adjusted based on the ALR as variables, using the y_1 variable previously explained as a dependent variable. In our study, to compute the ALR's, we've decided to take cluster 3 as a reference. Taking the most frequent cluster as a reference makes sense because it provides a representative baseline and it provides numerical strength: lower risk of divisions by small values, often lower variance of coefficients, and placing it in the denominator does not make it disappear, but rather it continues to enter the model implicitly; furthermore, it very often matches with the most solvent cluster.

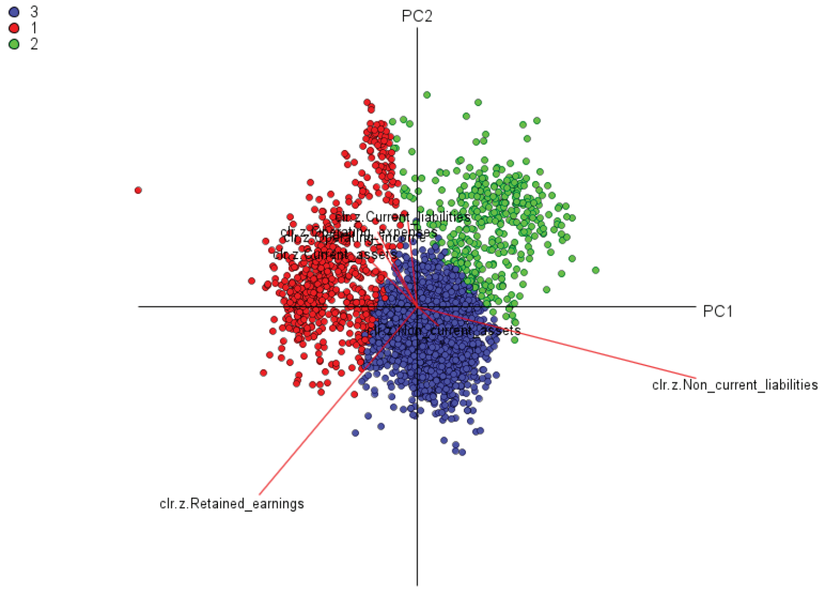

FIGURE 2
CoDa Biplot corresponding to a 3-cluster solution

Clus_1	Clus_2	Clus_3
0,77225396	0,03168918	0,19605686
0,25618493	0,04129127	0,7025238
0,31176553	0,50931951	0,17891496
0,81824971	0,06819871	0,11355158
0,34299634	0,04488858	0,61211508
0,9439142	0,01422682	0,04185898
0,64662011	0,11726004	0,23611985
0,52818734	0,32155389	0,15025877
0,18200329	0,67862531	0,1393714
0,93629248	0,02145508	0,04225244

TABLE 3
Membership values obtained for the first ten firms selected

The estimated results for 3151 firms were as follows:

$$\operatorname{logit}(y_1) = -3.16685 + 0.01257 \cdot alr_1 + 0.56484 \cdot alr_2 \tag{8}$$

```
Console
> >> m < glm(Bankruptcy ~ alr1 + alr2, data = Membership_v, family = binomial(
    family = binomial(link =  "logit")
Call:
glm(formula = Bankruptcy ~ alr1 + alr2, family = binomial(link ="logit"),
     data = Membership_v)

Deviance Residuals:
                IQ    Median      3Q       Max
(Intercept).0231   -0.20118  -0.1439     3.0695

Coefficients:
             Estimate  Std. Error  z value Pr(>|z|)
(Intercept) -3.16685     0.11661  -27.158   -27.158 ***
alr1         0.01257     0.04797    0.262     0.793
alr2         0.56484     0.06535    8.644   <2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 1 ' 1

Dispersion parameter for binomial family taken to be 1

    Null deviance: 947.56 on 3150 degrees of freedom
    Residual deviance: 841.59 on 3148 degrees of freedom
    AIC: 847.59

    Number of Fisher Scoring iterations: 7
```

FIGURE 3
Logit model implemented in R

- The coefficient $\beta_2 = 0.56484$ is positive and highly significant ($p < 2 \cdot 10^{-16}$), the odd-ratio for an increase of one unit in alr_2 is $e^{0.56484} \approx 1.76$. Therefore, as the logarithm is defined in base 2, an increase of one unit in alr_2 is equivalent to its classic ratio almost doubling. Therefore, if the ratio of cluster 2 doubles, the probabilities of bankruptcy multiply by approximately 1.76, keeping constant the rest.
- The coefficient $\beta_1 = 0.01257$ is positive but no significant ($p = 0{,}793$), the odd-ratio for alr_1 is $e^{0.01257} \approx 1.013$.

The expected value of the dependent variable (y1) increases when increasing the relative importance of components with positive coefficients in the equation, especially those with the largest coefficients in absolute value, at the expense of reducing that of components with negative coefficients. So in our case, belonging to cluster 2 carries the highest probability to be in a serious risk of business bankruptcy.

4 CONCLUSIONS

This study proposes and empirically evaluates an integrated methodological framework that combines Compositional Data Analysis (CoDa) and fuzzy cluster analysis to improve bankruptcy prediction in the agricultural retail sector. By explicitly addressing the relative nature of financial statements and overcoming the well-documented limitations of traditional financial ratios—such as asymmetry, redundancy, and sensitivity to outliers—the proposed approach provides a statistically coherent and economically meaningful alternative for financial distress assessment [13], [16] and [19].

The results show that transforming financial statement data into centered log-ratios prior to clustering yields well-defined financial profiles that are not distorted by scale effects or extreme observations. This finding is consistent with previous research demonstrating that CoDa transformations lead to more symmetric distributions and more reliable distance structures than conventional ratio-based approaches [8], [29] and [32]. In contrast to classical financial ratio clustering, the compositional approach avoids the formation of spurious or singleton clusters, a problem widely documented in the literature [13], [21].

The application of fuzzy clustering further enhances the analysis by allowing firms to display partial membership across multiple financial profiles. This non-exclusive classification is particularly appropriate in financial contexts, where company characteristics often evolve gradually and rarely conform to sharply defined categories [37], [40] and [42]. The fuzzy membership values capture transitional and hybrid financial situations that would be obscured under hard clustering schemes, as previously highlighted [9], [11] in financial applications of fuzzy classification.

A central contribution of this study lies in the treatment of fuzzy cluster membership values as compositional data and their transformation using additive log-ratios before inclusion in a logistic regression model. This step ensures full consistency with compositional data theory and avoids the interpretational and inferential problems that arise when compositional predictors are used in their raw form [29], [38]. The resulting logit model links bankruptcy probability to firms' relative affiliation with different financial profiles, allowing regression coefficients to be interpreted in terms of relative, rather than absolute, financial positioning.

From a practical perspective, the combined CoDa–fuzzy clustering–logit framework provides analysts, managers, and policymakers with a flexible tool for early warning systems. The probabilistic nature of the logit model, together with the non-exclusive classification of firms, enables nuanced risk assessments that better reflect real-world financial dynamics. Moreover, because the methodology is grounded in the compositional structure of financial statements rather than in sector-specific ratios, it is readily transferable to other industries and institutional contexts [8], [39].

Compared to the earlier study published in this journal, which focused on the methodological inadequacy of classical financial ratios for firm classification and proposed a CoDa–fuzzy clustering framework as a coherent alternative [9], the present work represents a substantive methodological and empirical extension. While the previous article demonstrated that combining compositional log-ratios with fuzzy clustering yields more coherent and stable classifications of firms' financial structures, its scope was limited to unsupervised classification and methodological validation [9], [11], [29]. Finally, this article extends the applicability of the CoDa–fuzzy clustering framework to a concrete decision-making context within the agricultural retail sector, using recent firm-level data and an explicit definition of bankruptcy grounded in both accounting and legal criteria. In doing so, it complements the earlier methodological contribution by demonstrating the practical relevance of fuzzy compositional classifications for financial risk assessment and early warning systems, an aspect only implicitly suggested in the previous work.

Several limitations should be acknowledged. The analysis focuses on a single sector and a specific economic environment, which may limit the generalizability of the findings. In addition, bankruptcy is defined using accounting and legal criteria that, while standard, may not capture all forms of financial distress. Future research could extend this framework by incorporating longitudinal analyses, alternative definitions of distress, or additional compositional predictors, including ESG-related financial components, as suggested in recent CoDa accounting studies [6].

ACKNOWLEDGEMENT

This research was supported by the Spanish Ministry of Science and Innovation /AEI/10.13039/501100011033 and by ERDF A way of making Europe (grant PID2021-123833OB-I00).

REFERENCES

[1] Moreno-Pérez, O.M.,& Blázquez-Soriano, A. (2023). What future for organic farming? Foresight for a smallholder Mediterranean agricultural system. *Agricultural and Food Economics* 11 (34).
[2] Hernandez-Romero, M., & Coenders, G. (in press). Financial resilience of agricultural and food production companies in Spain: A compositional cluster analysis of the impact of the Ukraine-Russia war (2021-2023). *European Accounting and Management Review. https://doi.org/10.48550/arXiv.2504.05912*
[3] Altman, E. I. (1968). Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy. *Journal of Finance*, 23(4), 589-609.
[4] Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18(1), 109-131.
[5] Arimany-Serrat, N., & Coenders, G. (2025). Biodiversidad y contabilidad: Metodología composicional en el análisis contable del sector apícola. *Economía Agraria y Recursos Naturales*, 25(1), 5-34.
[6] Arimany-Serrat, N., & Sgorla, A. F. (2024). Financial and ESG analysis of the beer sector pre-and post-COVID-19 in Italy and Spain. *Sustainability*, 16(17), 7412.
[7] Saus-Sala, E., Farreras-Noguer, M.À., Arimany-Serrat, N., & Coenders, G. (2024). Financial analysis of rural tourism in Catalonia and Galicia pre- and post COVID-19. *International Journal of Tourism Research*, 26(4), article e2698.

[8] Coenders, G., & Arimany-Serrat, N. (2023). Accounting statement analysis at industry level. A gentle introduction to the compositional approach. arXiv, 2305.16842.

[9] Linares-Mustarós, S., Coenders, G., & Vives-Mestres, M. (2018). Financial performance and distress profiles. From classification according to financial ratios to compositional classification. *Advances in Accounting*, 40, 1-10.

[10] Sala, E. S., Serrat, N. A., & Coenders, G. (2023). Análisis de las empresas de turismo rural en Cataluña y Galicia: rentabilidad económica y solvencia 2014–2018. *Cuadernos del CIMBAGE*, 1(25), 33-54.

[11] Molas-Colomer, X., Linares-Mustarós, S., Farreras-Noguer, M.À., & Ferrer-Comalat, J.C. (2024). A new methodological proposal for classifying firms according to the similarity of their financial structures based on combining compositional data with fuzzy clustering. *Journal of Multiple-Valued Logic and Soft Computing*, 43(1-2), 73-100.

[12] Magrini, A. (2025). Bankruptcy risk prediction: A new approach based on compositional analysis of financial statements. *Big Data Research*, 41, 1000537.

[13] Frecka, T. J., & Hopwood, W.S. (1983). The effects of outliers on the cross-sectional distributional properties of financial ratios. *Accounting Review*, 58(1), 115-128.

[14] Deakin, E. B. (1976). Distributions of financial accounting ratios: Some empirical evidence. *The Accounting Review* 51(1), 90-96.

[15] Ezzamel, M., & Mar-Molinero, C. (1990). The distributional properties of financial ratios in UK manufacturing companies. *Journal of Business Finance & Accounting*, 17(1), 1-29.

[16] McLeay, S., & Omar, A. (2000). The sensitivity of prediction models to the non-normality of bounded and unbounded financial ratios. *The British Accounting Review*, 32(2), 213-230.

[17] Iotti, M., Manghi, E., & Bonazzi, G. (2023). Financial performance of companies associated with the PDO Parma ham consortium: Analysis by quartile of firms. *Journal of Agriculture and Food Research*, 13, 100598.

[18] Durana, P., Kovalova, E., Blazek, R., & Bicanovska, K. (2025). Business efficiency: Insights from Visegrad four before, during, and after the COVID-19 pandemic. *Economies*, 13(2), 26.

[19] Lev, B., & Sunder, S. (1979). Methodological issues in the use of financial ratios. *Journal of Accounting and Economics*, 1(3), 187-210.

[20] Cowen, S. S., & Hoffer, J. A. (1982). Usefulness of financial ratios in a single industry. *Journal of Business Research*, 10(1), 103-118.

[21] Watson, C.J. (1990). Multivariate distributional properties, outliers, and transformation of financial ratios. *The Accounting Review*, 65(3), 682-695.

[22] Creixans-Tenas, J., Coenders, G., & Arimany-Serrat, N. (2019). Corporate social responsibility and financial profile of Spanish private hospitals. *Heliyon* 5 (10), e02623.

[23] Feranecová, A., & Krigovská, A. (2016). Measuring the performance of universities through cluster analysis and the use of financial ratio indexes. *Economics & Sociology*, 9(4), 259-271.

[24] Santis, P., Albuquerque, A., & Lizarelli, F. (2016). Do sustainable companies have a better financial performance? A study on Brazilian public companies. *Journal of Cleaner Production*, 133, 735-745.

[25] Sharma, S., Shebalkov, M. & Yukhanaev, A. (2016). Evaluating banks performance using key financial indicators – a quantitative modeling of Russian banks. *The Journal of developing Areas*, 50(1), 425-453.

[26] Chen, K. H., & Shimerda, T. A. (1981). An empirical analysis of useful financial ratios, *Financial Management* 10 (1), 51-60.

[27] Pohlman, R.A., & Hollinger, R.D. (1981). Information redundancy in sets of financial ratios. *Journal of Business Finance & Accounting*, 8(4), 511-528.

[28] Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B* (Methodological), 44(2), 139-160.

[29] Aitchison, J. (1986). The statistical analysis of compositional data. *Monographs on Statistics and Applied Probability.* London: Chapman and Hall.

[30] Pawlowsky-Glahn, V., & Buccianti, A. (2011). Compositional data analysis. *Theory and applications*. Wiley, New York.

[31] Van den Boogaart, K.G., & Tolosana-Delgado, R. (2013). Analyzing compositional data with R. *Springer*. Berlin.

[32] Pawlowsky-Glahn, V., Egozcue, J. J., & Tolosana-Delgado R. (2015). Modeling and analysis of compositional data. Wiley.

[33] Filzmoser, P., Hron, K., & Templ, M. (2018). Applied compositional data analysis with worked examples in R, *Springer*, New York.

[34] Greenacre, M. (2018). Compositional data analysis in practice. *Chapman and Hall/CRC press*, New York.

[35] Bezdek, J. C., & Pal, S. (1992). Fuzzy Models for Pattern Recognition: Methods that Search for Structures in Data. *IEEE Press*.

[36] Yang, M.S. (1993). A survey of fuzzy clustering. *Mathematical and Computer modeling*, 18(11), 1-16.

[37] Höppner, F., Klawonn, F., Kruse, R. & Runkler, T. (1999). Fuzzy cluster analysis: methods for classification, data analysis and image recognition. *John Wiley & Sons*.

[38] Coenders, G., & Pawlowsky-Glahn, V. (2020). On interpretations of tests and effect sizes in regression models with a compositional predictor. *SORT-Statistics and Operations Research Transactions*, 44(1), 201-220.

[39] Arimany-Serrat, N., Farreras-Noguer, M. À., & Coenders, G. (2023). Financial resilience of Spanish wineries during the COVID-19 lockdown. *International Journal of Wine Business Research*, 35(2), 346–364.

[40] Zadeh L.A. (1965). Fuzzy sets. *Information and Control* 8, 338-353.

[41] Dunn, J. C. (1973). A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters, *Journal of Cybernetics* 3: 32-57.

[42] Bezdek, J.C. (1981), Pattern Recognition with Fuzzy Objective Function Algorithm, *Plenum*, NY.

[43] Kaufman, L., & Rousseeuw, P. J. (1990). Finding groups in data: An introduction to cluster analysis. *John Wiley & Sons*.

[44] Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1), 1-27.