An empirical examination of financial performance and distress profiles during Covid-19: the case of fishery and food production firms in Vietnam

Journal:	*Journal of Financial Reporting and Accounting*
Manuscript ID	JFRA-09-2023-0509.R3
Manuscript Type:	Research Paper
Keywords:	Weighted compositional data analysis, accounting ratios, fishery and food industries, cluster analysis, Covid-19, Vietnam

Article Title Page

An empirical examination of financial performance and distress profiles during Covid-19: the case of fishery and food production firms in Vietnam

Author Details:

Binh Thi Thanh Dao
Faculty of Management and Tourism, Hanoi University, Hanoi, 100000, Vietnam

Germà Coenders
Department of Economics, Faculty of Economics and Business, University of Girona, Campus Montilivi, 17003 Girona, Spain

Phuong Lai Hoai
Faculty of Management and Tourism, Hanoi University, Hanoi, 100000, Vietnam

Trang Thi Thu Dam
Faculty of Mathematical Economics, Thuongmai University, Hanoi, 100000, Vietnam

Huong Thi Trinh*
Faculty of Mathematical Economics, Thuongmai University, Hanoi, 100000, Vietnam

Corresponding author: Huong Thi Trinh
[Corresponding Author's Email]: trinhthihuong@tmu.edu.vn

Acknowledgments (if applicable): The authors would like to thank the reviewers and the editor for their helpful comments and suggestions which improved the presentation of the paper. Part of this work was completed while the last author was visiting the Vietnam Institute for Advanced Study in Mathematics (VIASM). The author would like to thank the Institute for its hospitality and support. This paper has also benefited from valuable discussions with participants at Vietnamese Economists Annual Meeting, Danang, July 2023, 10th National Mathematical Congress, Danang, August 2023, Vietnam Symposium in Supply Chain Management (VSSCM-2023), Hanoi, October 2023, 4th Annual meeting of Vietnam Applied Statistics Network (VASN), Hanoi, April 2024. Finally, we would like to express our gratitude to Christine Thomas-Agnan for proofreading the English text.

Funding
This research is supported by Ministry of Education and Training of Vietnam (Grant number: B2024-NHF-04), the Spanish Ministry of Science and Innovation /AEI/10.13039/501100011033 and by ERDF – A way of making Europe [grant number PID2021-123833OB-I00], the Department of Research and Universities of the Generalitat de Catalunya [grant numbers 2021SGR01197 and 2023-CLIMA-00037] and the Spanish Ministry of Health [grant number CIBERB06/02/1002].

An empirical examination of financial performance and distress profiles during Covid-19: the case of fishery and food production firms in Vietnam

Abstract

Purpose - Financial ratios are often utilized to classify firms into different clusters of financial performance. This study aims to classify firms using financial ratios with advanced techniques and identify the transition matrix of firms moving clusters during the Covid-19 period.

Design/methodology/approach - This article employs compositional data (CoDa) analysis based on existing clustering methods with transformed data by weighted logarithms of financial ratios. The data include 66 listed firms in Vietnam's food and beverage and fishery sectors over a three-year period from 2019 to 2021, including the Covid-19 period.

Findings - These firms can be classified into three clusters of distinctive characteristics, which can serve as benchmarks for solvency and profitability. The results also show the migration from one cluster to another during the Covid-19 pandemic, allowing for the calculation of the transition probability or the transition matrix.

Practical implications – The findings indicate three distinct clusters (good, average, and below-average firm performance) that can help financial analysts, accountants, investors, and other strategic decision-makers in making informed choices.

Originality – Clustering firms with their financial ratios often suffers from various limitations, such as ratio choices, skewed distributions, outliers, and redundancy. This study is motivated by a weighted CoDa approach that addresses these issues. This method can be extended to classify firms in multiple sectors or in other emerging markets.

Keywords: Weighted compositional data analysis, accounting ratios, fishery and food industries, cluster analysis, Covid-19, Vietnam.

1. Introduction

The World Health Organisation (WHO) declared the outbreak of a novel coronavirus (Covid-19) pandemic in March 2020. In Vietnam, this infectious disease had resulted in over 11.5 million confirmed cases and more than 43,000 deaths by 31 March 2023 (WHO, 2023). During both 2020 and 2021, the country tightened rules to contain the spread of the disease, dictated social distancing measures (including quarantine), and various partial and national lockdowns were imposed, leading to a significant decline in demand for products, a severe shortage of inputs for production, a drastic decline in investment and disappointed expectations. As a result, the financial health of local firms suffered severely. Given its unique speed of change and scale of impact on all segments of the economy, the consequences of Covid-19 are worth examining (Boubaker *et al.*, 2022). There have been numerous studies on the impact of the Covid-19 crisis on firm performance (see Boubaker *et al.*, 2023 for a systematic review); however, there is an additional avenue to explore: the downturn/upturn in performance of individual companies or how individual firms navigate the economic disruptions due to the spread of the disease and the lockdowns (Kao *et al.*, 2023). Therefore, it is crucial to classify firms in terms of their financial positions and their deterioration/improvement over time.

Firm classification or clustering is the technique that groups heterogeneous firms into meaningful homogeneous groups. The method intends to form groups with the highest possible internal cohesion and external separation (Capece *et al.*, 2010). Its usefulness has resulted in the use of this method in various contexts and disciplines (Caruso *et al.*, 2018). For example in the management field, it can help to study entrepreneurial characteristics (Parkes *et al.*, 2018). In finance, clustering or classification algorithms have been applied in credit scoring to differentiate borrowers by riskiness (Caruso *et al.*, 2021; Williams and Huang, 1997; Yeo *et al.*, 2001). In business, classification has also been used as a tool to analyze firm performance (RL and Mishra, 2022; Rodrigues and Rodrigues, 2018; Toly *et al.*, 2020), which can help to identify clusters of firms with superior and inferior financial performance. A clear picture of the firms' health can assist investors in making informed decisions about their investments and how to allocate resources. Firm owners and managers can compare their firm's performance to their closest cluster rather than to the global average of an industry that may not be homogeneous. Firm classification can also improve stock price prediction and trading returns (Sáenz *et al.*, 2023). Financial ratios are regarded as fundamental instruments in financial analysis, and their use has been proven to provide accurate and complete assessments of firm performance (Krylov, 2018; Saleh *et al.*, 2023). This paper thus examines the use of financial ratios in firm classification.

While standard financial ratios can be helpful in diagnosing individual firms, they have been found to exhibit some serious statistical and practical problems when used as variables in statistical analyses within one or several industries. These problems include asymmetry, non-linearity, outliers, severe non-normality, mutual redundancy in ratios that measure overlapping concepts, and even dependence on arbitrary decisions regarding which accounting figure is placed in the numerator and which in the denominator (Carreras-Simó and Coenders, 2021; Cowen and Hoffer, 1982; Creixans-Tenas *et al.*, 2019; Faello, 2015). In the case of cluster analysis, these problems lead to the results failing to capture distinct financial profiles and to clusters composed only by outliers (Jofre-Campuzano and Coenders, 2022; Linares-Mustarós *et al.*, 2018). These authors suggest the use of the Compositional Data (CoDa) methodology as a more valid alternative.

Financial ratios constitute a genuine case of researchers' and professionals' interest in relative rather than absolute magnitudes and thus a natural field of application of CoDa, which has the same objective (Aitchison, 1986; Coenders *et al.*, 2023). Essentially, a composition is an array of strictly positive numbers for which only the relative information between them is of interest (Egozcue and Pawlowsky-Glahn, 2019), which perfectly fits the notion of financial ratios. The CoDa methodology offers several advantages over standard financial ratios in the statistical analysis of ratios. CoDa treat accounting figures symmetrically so that results do not depend on numerator and denominator permutation. The CoDa method also reduces outliers, redundancy, non-linearity, and non-normality. Numerous studies have explored the applications of CoDa in accounting and financial statement analysis (Arimany-Serrat *et al.*, 2022, 2023; Carreras-Simó and Coenders, 2021; Creixans-Tenas *et al.*, 2019; Linares-Mustarós *et al.*, 2022). There are also several recent studies on financial profile classification using the CoDa approach (Jofre-Campuzano and Coenders, 2022; Linares-Mustarós *et al.,* 2018; Saus-Sala *et al.*, 2021). Far from being a statistical refinement, the CoDa approach leads to very substantial differences in the analysis results whenever they are compared with standard financial ratios (Arimany-Serrat *et al.*, 2022; Carreras-Simó and Coenders, 2021; Jofre-Campuzano and Coenders, 2022; Linares-Mustarós *et al.*, 2018, 2022). Saus-Sala *et al.* (2021) identified three distinct clusters of farm-tourism firms based on financial profiles, including leverage, margin, and turnover ratios. Similarly, Jofre-Campuzano and Coenders (2022) applied weighted CoDa in clustering companies of automotive fuel in Spain and found four clusters, one of which characterized by financial distress. Likewise, Linares Mustarós *et al.* (2018) found three clusters in the pharmaceutical industry and four clusters in the wearing apparel industry. The above studies demonstrate new insights on clustering firms' financial health and its potential application across diverse industries and economic contexts.

This article focuses on two key sectors of Vietnam's economy, namely food and beverage (FB), and fishery, which contribute significantly to the country's gross domestic product and were among those severely affected by the pandemic. The former added USD 17 billion to Vietnam's GDP and employed three million people in 2021. Meanwhile, the figures for the latter were USD 11 billion and four million people, respectively. According to MFAT (2022), retail sales of FB in Vietnam rose by roughly 10% per annum from 2016 to 2020 and reached USD 51 billion in 2020. However, this industry is among the hardest hit by the pandemic. According to a report by HouseLink (Nguyen *et al.*, 2021), the index of industrial production (IIP) of the food industry boasted an average annual growth of over 7% prior to the crisis. However, the figure only grew by 5.3% in 2020 and 1.1% in the first 10 months of 2021. Meanwhile, the beverage sector could not maintain a similar growth rate as its IIP dropped by 5.2% and 5.8% in 2020 and the first 10 months of 2021, respectively. In 2021, total sales of FB listed firms reached VND 279,000 billion (up 5.6% from the same period) (Nguyen, 2022). However, gross profits only rose by 1.1% year on year, due to a decline in gross profit margin and an increase in the cost of raw materials. In addition, these businesses' selling and administrative expenses were on the rise due to the pandemic, which required businesses to increase spending on disinfection measures and promotional programs. As a result, operating profits decreased by 2.3% to VND 33,935 billion. Meanwhile, statistics from the Southeast Asia Fisheries Development Center show that Vietnam was ranked in the top ten fishery and aquaculture producers before the pandemic and was a major exporting country. Fishery constitutes a crucial sector, particularly in terms of poverty alleviation, economic growth, and employment provision (SEAFDEC, 2022). Nonetheless, the pandemic instantly hit the industry, leading to a drop in exports in 2019 and 2020, before a gradual recovery started in 2021 (VASEP, 2022). The FB and fishery industries are vital contributors to people's daily lives and play an essential role in Vietnam's economy. They are worth examining in the context of Covid-19.

In this article, we follow Jofre-Campuzano and Coenders (2022) using weighted CoDa cluster analysis and complement it with transition matrices to capture the dynamics in cluster membership attributable to the pandemic. The main objectives of this article are thus to (1) construct a classification of Vietnamese firms in these two industries that allows each firm to compare itself with the cluster of the closest firms to itself, and (2) identify the financial ratios that contribute to the transition of these firms into different performance clusters between 2019 and 2021, including some of the most common financial ratios related to solvency and profitability. Weighted CoDa analysis helps to improve the accuracy and robustness of classification. This method is simple and yet has high potential for application in other sectors and emerging countries to provide insights for financial analysts, accountants, investors, and researchers.

The remainder of the article is organized as follows: Section 2 presents the hypotheses; Section 3 outlines the method that we use for firm classification; Section 4 presents the empirical results; Section 5 provides a discussion of the results; and Section 6 summarizes and concludes.

2. Hypothesis development

The hypotheses for this study are developed below.

Apedo-Amah *et al.* (2020) shed light on the firm-level heterogeneity in the impact of and response to Covid-19, suggesting that firms are indeed different in their reaction to the shocks caused by the pandemic (Brancati and Brancati, 2020). Fernández-Cerezo *et al.* (2023) and Liu *et al.* (2023) share the same findings, confirming heterogeneous effects of the pandemic across industries and firms of different characteristics. In addition, some drivers of differential financial resilience to Covid-19 are identified by Arimany-Serrat *et al.* (2023) and Boubaker *et al.* (2022). This leads to the following hypothesis:

H_1: The FB and fishery industries are composed of distinct clusters with different performance during the Covid-19 pandemic.

Sharif et al. (2020) demonstrated that the heightened uncertainty during the epidemic would postpone the financial choices made by businesses and diminish social spending, debt issuance, and investment activities. The uncertain effects of policies and regulations also influence commodities, including basic goods such as oil and gasoline (Sharif *et al.*, 2020). It has been indicated that the crisis brought about by the pandemic has impacted business operations in terms of capital shortages (Halling *et al.*, 2020, turnover and margin (Arimany-Serrat *et al.*, 2023), supply chain operations (Sharma *et al.*, 2020), and labor markets (Jordà *et al.*, 2022), and heightened the risk of default (Ho *et al.*, 2023). This leads to the following hypothesis:

H_2: Firms tend to move to worse performing clusters during the pandemic.

Conversely, Kong *et al.* (2022) reported that micro and small Chinese firms displayed a notable improvement in performance one year following the onset of the Covid-19 pandemic. Cirera *et al.* (2021) also mentioned some recovery since the pandemic. This leads to the following hypothesis:

H_3: Firms tend move back to better preforming clusters after the pandemic.

There has been intensive empirical research indicating that better solvency and higher profitability are associated with firms of better performance (Noviyana *et al.*, 2023; Chue and Xu, 2022; Dahiyat *et al.*, 2021). This leads to the following hypothesis:

H_4: Clusters with better profitability also have better solvency.

3. Method

This section briefly describes the CoDa method and the weighted CoDa method for financial statement analysis, how to transform financial statements with the CoDa method, and the analysis steps for compositional cluster analysis of financial-statement data. More detail about the CoDa method is provided in Appendix 1.

3.1. Compositional financial statement analysis

Compositions are arrays of D positive figures, called *parts*, whose values are of interest in relative terms (Aitchison, 1986):

$$(x_1, x_2, \ldots, x_D) \text{ with } x_j > 0, j = 1,2, \ldots, D. \quad (1)$$

In financial statement analysis, the parts are accounting figures resulting from the financial statements. In order to introduce accounting statements in a D-part composition negative figures have to be avoided (Creixans-Tenas *et al.*, 2019). Even though financial ratios sometimes involve accounting figures that may be negative, their use is advised against in the financial literature because they can cause discontinuity, outliers, or even a reversal of interpretation when the accounting figure which may be negative is in the denominator (Lev and Sunder, 1979). Besides, computing a ratio is a meaningful operation only for variables in a ratio scale, which require a meaningful absolute zero and thus cannot have negative values (Stevens, 1946).

Accounting figures are generally negative because they imply some form of subtraction of other positive accounting figures. For instance, one should directly use revenues and costs rather than profit or current assets and liabilities rather than working capital.

The ultimate choice of accounting figures to include will depend on the analysis objectives or the research questions. In principle, the researchers will select the accounting figures needed to compute their favorite financial ratios. In this article, the parts represented by the x_j parts are the following $D = 6$ positive financial statement account categories which have already been used in related research (Arimany-Serrat *et al.*, 2023; Creixans-Tenas *et al.*, 2019; Jofre-Campuzano and Coenders, 2022) :

$$\begin{aligned}
x_1 &= \text{Non} - \text{current assets} \\
x_2 &= \text{Current assets} \\
x_3 &= \text{Non} - \text{current liabilities (Long} - \text{term debt)} \qquad (2) \\
x_4 &= \text{Current liabilities} \\
x_5 &= \text{Revenues (net sales)} \\
x_6 &= \text{Costs (total costs)}
\end{aligned}$$

The standard financial ratios of interest, which can be computed from x_1 to x_6, include three ratios related to solvency and five related to the decomposition of profitability through DuPont analysis (Saus-Sala *et al.*, 2021; Soliman, 2008). Table 1 shows the definition of each ratio. Within solvency ratios, liquidity measures the ability to meet short-term payments, debt maturity indicates the part of debt whose payment can be postponed, and indebtedness is an overall solvency measure. Within profitability ratios, turnover measures the efficiency in the use of assets and margin measures the relative differences between revenues and total costs. Profitability itself can be defined with respect to assets or equity, with leverage being the conversion factor between the two.

Insert Table 1 about here

Following DuPont analysis, the following identities make it possible to decompose profitability (Houmes *et al.*, 2018):

$$\text{Return on Assets (ROA)} = \text{Turnover} \times \text{Margin}, \qquad (3)$$

$$\text{Return on Equity (ROE)} = \text{ROA} \times \text{Leverage}.$$

Thus, a high leverage can be a mixed blessing. When ROA is positive, it contributes to increasing ROE but reduces solvency. When ROA is negative, it even decreases ROE.

3.2. Weighted CLR coordinate computation

The usual approach to the CoDa method is to use existing standard statistical methods after transforming the CoDa into a Euclidean space (see Appendix 1). Then, traditional statistical methods can be applied for cluster analysis or any other purpose. Logarithms of ratios are the standard transformation in CoDa (Aitchison, 1986). The CoDa methodology ensures that *D* so-called centered logratios (CLR) contain all information about the relative importance of *D* accounting figures (Pawlowsky-Glahn *et al.*, 2015). An important property of the CLR transformation is that it preserves the distances. Precisely, the Aitchison distance between two compositions is equal to the usual Euclidean distance between the corresponding transformed vectors (see Appendix 1). Compared to standard financial ratios, log-ratios have fewer problems of non-linearity, asymmetry and outliers (Carreras-Simó and Coenders, 2021; Jofre-Campuzano and Coenders, 2022; Linares-Mustarós *et al.*, 2018, 2022).

Given *D* financial statement account categories (Equation (2)), the CLR transformation is defined as:

$$\mathrm{CLR_j} = \log\left(\frac{\mathrm{x_j}}{\mathrm{x}_1^{\frac{1}{D}}\mathrm{x}_2^{\frac{1}{D}}\cdots\,\mathrm{x}_D^{\frac{1}{D}}}\right) = \log\left(\frac{x_j}{\sqrt[D]{x_1\cdots x_D}}\right), \text{with } j = 1,\ldots,D. \qquad (4)$$

In the context of financial statement analysis, each CLR compares one accounting figure $\mathrm{x_j}$ with the geometric mean of all accounting figures for each firm (i.e, $\sqrt[D]{x_1\cdots x_D}$). The application of the CLR transformation in financial statement analysis is explored in the work of Saus-Sala *et al.* (2021). CLR can be used as raw data in multivariate descriptive analysis methods such as cluster analysis, for which purpose, their interpretation is not necessary, as demonstrated below.

In an uncritical application of the CoDa methodology it can happen that account categories which generally take low values have a very large CLR variance and a disproportionate influence on the classification results (Egozcue and Pawlowsky-Glahn, 2016; Greenacre, 2018; Greenacre and Lewi, 2009; Hron *et al.*, 2021, 2022; Jofre-Campuzano and Coenders, 2022). This limitation can be solved by computing a weighted counterpart to the CLR transformation in which the weights $\mathrm{w_j}$ adding up to 1 are proportional to the means of $\mathrm{x_j}$ (Greenacre and Lewi, 2009; Jofre-Campuzano and Coenders, 2022):

$$\mathrm{wCLR}_j = \sqrt{w_j}\log\left(\frac{x_j}{x_1^{w1}x_2^{w2}\cdots\,x_D^{wD}}\right), \text{ with } j = 1,\ldots,D. \qquad (5)$$

3.3. Weighted compositional cluster analysis steps

The steps we followed to conduct the weighted cluster analysis of financial statements draw from Jofre-Campuzano and Coenders (2022) and are presented in Figure 1. Prior to the computation of the weighted CLR, zero values have to be imputed (Martín-Fernández *et al.*, 2011). The compositional methodology offers an advanced toolbox for zero imputation under various assumptions (Martín-Fernández *et al.*, 2012), like the modified EM algorithm by Palarea-Albaladejo and Martín-Fernández (2008). For this purpose, we used the lrEM function in the R package zCompositions (Palarea-Albaladejo and Martín-Fernández, 2015). In our dataset, only 13.13% of non-current liabilities had zero values, a percentage that is suitable for replacement (Palarea-Albaladejo and Martín-Fernández, 2008). The remaining parts had no zeros.

In terms of implementing the classification, the CLR function in the R package easyCODA was used to compute the weighted CLR (Greenacre, 2018). Then, Ward clustering (Landau *et al.*, 2011; Ward Jr, 1963) was applied with the WARD function using the weighted CLR as input data. Ward clustering is a standard method for financial statement data among various clustering methods assuming Euclidean distances (Linares-Mustarós *et al.*, 2018). Scree plots of the within-cluster sums of squares were used to select the number of clusters (Dolnicar *et al.*, 2018).

The compositional center is defined as the array of the D geometric means of each accounting figure x_1 to x_D over the sample of firms (Pawlowsky-Glahn *et al.*, 2015), not to be mistaken with the geometric mean of each firm over all accounting figures used in Equation (4). For interpretation of the typical financial profile in each cluster, the standard financial ratios in Table 1 were computed using the compositional centers. In this way, interpretation can revert to the standard financial ratios (Saus-Sala *et al.*, 2021). The clusters were next related to non-financial variables with mosaic plots and confidence-interval plots (Dolnicar *et al.*, 2018; Goldstein and Healy, 1995).

Compared to Jofre-Campuzano and Coenders (2022), this article includes the additional step of assessing cluster dynamics by means of transition matrices.

Insert Figure 1 about here

In summary, step 1 is to ensure that CLR (as log) can be calculated. Step 2 deals with the calculation of input variables free of asymmetry and outliers and for which distance computation makes sense. Step 3 extracts the clusters. Step 4 profiles the representative ratios of each cluster. The results of these steps until step 5 are provided in Section 4.2. Steps 6 and 7 use the classification with respect to non-financial variables and the classification's own dynamics, and the results are discussed in Sections 4.3 and 4.4, respectively.

Utilizing weighted CoDa for firm classification represents an innovative approach. From just a few input variables extracted from firms' financial statements, the method can derive key financial metrics free of asymmetry and extreme outliers and effectively categorize firms.

3.4. Data

The study uses financial information extracted from the financial statements of all listed firms in the FB and fishery industries during a three-year period from 2019 to 2021. After excluding firms with missing values, the final sample consisted of 66 firms. Firms in both the FB and fishery industries produce necessities, and played a significant role during the times of the Covid-19 pandemic.

4. Data and results

4.1. Exploratory analysis

Table 2 shows the descriptive statistics of our dataset. We have data for three years and the same 66 companies each year. The FI (Fishery) industry accounts for 24.2% of the observations, and the remaining 75.8% belong to the FB industry.

The data relating to the six parts named x_1 to x_6, as explained in Equation (2) are presented with their means and the standard deviations in brackets. Most standard deviations are higher than the corresponding means, thus revealing substantial variability in the sample. Four of the nine financial ratios also have standard deviations higher than their means. The remaining five still have substantial variability relative to the corresponding means. This data characteristic hints at high industry heterogeneity and the need for classification. Other variables such as firm age, year of listing, ownership type (state-owned (SO), 39.4%, or private-owned (PO), 60.6 %), as well as the exchange where firms are listed, are also presented in Table 2. Most firms started being listed around 2010. The proportion of firms listed in the OTC (Over the Counter) and UPCOM (Unlisted Public Company Market) exchanges stands at 41. There is no variation in the exchange listing the companies between 2019 and 2021.

All analyses in this article, except the transition matrices, are based on the pooled dataset of 198 cases for three years.

Insert Table 2 about here

Figure 2A in Appendix 2 shows boxplots of standard financial ratios. Many ratios have outliers, particularly liquidity, debt maturity, turnover, leverage, and ROE. These ratios also exhibit asymmetry, especially liquidity, debt maturity, turnover, and leverage.

In Figure 3A in Appendix 2, the boxplots depict the weighted CLR. It is observed that there are some outliers near the whiskers, but no extreme outliers, unlike the standard financial ratios shown in Figure 2A. Following the weighted CLR, non-current liabilities have the most negative outliers. None of the CLRs display significant asymmetry.

After zero imputation, we calculate the arithmetic mean for each part. Then, the arithmetic means are normalized to have a unit sum (Table 3). These mean values are used as weights for Equation (5) (steps 1 and 2 of Figure 1). As in Jofre-Campuzano and Coenders (2022), the non-current liabilities have the lowest mean by far. Thus, the weighted method can help reduce the extreme contribution to the total variance of non-current liabilities and its outliers, leading to a better-balanced contribution of the parts (Table 4). Especially, the contribution of non-current assets and current liabilities has improved, compared to their very poor representation in an unweighted analysis.

Insert Table 3 about here

Insert Table 4 about here

4.2. Weighted classification

Figure 4 presents the dendrogram and the scree plot resulting from the Ward cluster analysis (Step 3 in Figure 1.) The left figure shows an obvious 3-group solution if the cutting point is between 0.15 and 0.30. Similarly, the right figure also shows the elbow cut at three clusters.

Insert Figure 4 about here

Table 5 shows the typical financial profile in each cluster as the standard financial ratios computed from the cluster geometric means of x_1 to x_6. Clusters are ordered from the highest financial performance to the highest financial distress (Steps 4 and 5 in Figure 1).

It is very clearly seen that Cluster 1 represents the best-performing firms and Cluster 3 the worst ones in the sample. In Cluster 1, the performance measures, namely ROA, ROE, margin, and turnover, have the highest means. In contrast, these ratios are lowest in Cluster 3. Indebtedness and leverage have the highest means in Cluster 3, which also has the worst liquidity, thus representing a potential financial distress group. Cluster 1 represents the worst debt maturity, but this is irrelevant when indebtedness is low, and liquidity is high. Cluster 2 has slightly worse turnover, liquidity, and indebtedness than Cluster 1, the main difference being a substantially lower margin, which translates into lower ROA and ROE, which are still acceptable. By the number of cases, the distressed Cluster 3 is the largest and Cluster 1 is the smallest. The non-distressed Clusters 1 and 2 together account for just 41% of cases. The analysis provides distinct and informative clusters which can help to establish benchmarks for the classification of companies based on solvency and profitability.

Insert Table 5 about here

4.3. Relationship between the clusters and external variables

This section presents the relationship of cluster membership to selected qualitative and numeric variables using mosaic plots and confidence-interval plots (step 6 in Figure 1).

In the mosaic plots in Figure 5, the bar widths indicate the frequency distribution by categories of industry, type of stock exchange and year of listing. The bar heights indicate the proportion of firms in each cluster within each category. The proportion of companies in Cluster 1 (resp. Cluster 2) is higher (resp. lower) for the Fishery (FI) industry than in the FB industry. The proportion in Cluster 3 is almost the same for both industries. The HNX exchange has the highest proportion of companies belonging to Clusters 1 and 2. Meanwhile, UPCOM boasts the largest number of firms among the four stock exchanges and the second-highest proportion of cluster 1-graded companies, only after HNX. The HOSE exchange stands out for Cluster 3. The OTC exchange has only three cases, all in Cluster 2. Firms listed before 2006 are the most frequent in Cluster 3, while none belong to Cluster 1. Firms listed between 2006 and 2010 stand out for Clusters 1 and 2. Firms listed in 2016 and after are the second most related to Cluster 3. The proportions belonging to Cluster 1, 2, or 3 for private-owned or state-owned companies are almost the same (data not shown).

Insert Figure 5 about here

Following Goldstein and Healy (1995), when two intervals drawn with 83% confidence do not overlap, there is a significant difference at 95% confidence between the corresponding two groups of companies with respect to the variable on the vertical axis. Figure 6 shows that the clusters significantly differ in total assets and revenues, which are indicators of company size. Cluster 1 contains the smallest companies according to both criteria. Meanwhile, company age is not related to the classification of firms (data not shown).

Insert Figure 6 about here

4.4. Transition matrix analysis

Regarding step 7 in Figure 1, Table 6 shows that during the containment year (2020), five firms were re-categorized into groups with inferior performance. Four of these firms were from the FB industry. Specifically, one FB company was downgraded from Cluster 1 to Cluster 3. A detailed examination of this firm showed that the possible reasons for the downward movement were significant increases in debt maturity (from 0 to 0.199) and in indebtedness (by 50%). A drop in income (by 46% in ROA and by 32.2% in ROE) may also have contributed to this transition. Another FB firm in Cluster 1 moved down to Cluster 2, primarily due to a 13% decrease in ROE.

Meanwhile, three other firms transitioned from Cluster 2 to the worst cluster (Cluster 3). Possible reasons for the transition include a significant drop in profitability (ROA, ROE, margin) or a deterioration in the liquidity position. It can be deduced that if a firm witnessed a weakening performance in any of the criteria, it faced the possibility of being downgraded. Table 6 is thus a good summary of the pandemic effect. Having said this, 61 of the 66 firms stayed in the same cluster between 2019 and 2020.

Insert Table 6 about here

Insert Table 7 about here

As seen in Table 7, two FB firms managed to make a comeback from Cluster 3 to Cluster 2. Both upgrades were related to a decline in the debt position (debt maturity, indebtedness, and leverage all moved downward) or an improvement in liquidity and turnover. Once again, this implies that the chance of being upgraded may depend on the progression of certain ratios, not all. More importantly, it implies that, for some firms, recovery after the pandemic was fast.

5. Discussion

Any industry is diverse according to heterogeneous strategic choices regarding margin, turnover, leverage, size, ownership structure and so on. In this respect, cluster analysis, by identifying more than one average financial profile, is more realistic and helpful in providing financial benchmarks. The analysis shows that the three clusters have markedly different profiles. Specifically, Cluster 3 is the least profitable (worst margin, leverage, ROA, and ROE), and the most indebted (worst indebtedness, leverage, and liquidity, albeit with the most extended debt maturity). Meanwhile, Cluster 1 boasts the highest profitability, lowest debt burden, and the best turnover. Cluster 2 is somewhere between these two groups, with lower profitability than Cluster 1 but about equally good indebtedness and liquidity. Most companies are categorized into Cluster 3, while few are grouped into Cluster 1.

Since the sampled firms are from two different sectors, we considered both industry and cluster. Within a sample of 16 fisheries, 19% belong to the best cluster, another 19% belong to the second cluster, and as many as 62% to cluster 3. Among the sample of 50 food-and-beverage firms, the proportion of the poorest performers is the same as among the fisheries (62%). However, the percentages of firms in Clusters 1 and 2 differ, with 10% and 28%, respectively. Thus, fisheries have a slightly better performance.

Confidence-interval analysis (Figure 6) confirms that firm characteristics appear to influence the classification. In terms of revenues or assets, small firms belong to the cluster with the best financial health, while large companies are in vulnerable states. The mosaic plots show that a higher proportion of firms in poor financial health is associated with being listed before 2006 or after 2016.

The transition matrices (Tables 6 and Table 7) show that 61 out of 66 firms stayed within their clusters without transitioning. The results imply some stability in the performance of these firms even during the 2020 containment year, thus ensuring a meaningful classification of businesses in the sample over time. Besides, it can be argued that since the two industries, FB and fishery, produce necessary goods, the blow that the pandemic dealt to these sectors was not fatal. However, our analysis shows that Covid-19 still exerts an immediate effect on the performance of these firms, indicated by the transition of five out of 66 firms (8%) to worse clusters in 2020. This transition coincides with increased public fears, an imposition of travel bans, restrictions on economic activity, and a rise in transaction costs. For example, the FB industry was strongly hit by the mass closure of food establishments, driven by public fears and travel restrictions or lockdowns (VIRAC, 2021). This resulted in lower domestic demand and challenges for FB producers and service providers. Additionally, disruptions in imports and exports due to the Covid-19 pandemic posed difficulties for suppliers and traders in distributing, and maintaining sufficient inventory of goods, leading to increased costs. Meanwhile, according to To *et al.* (2023), the fishery industry suffered from diminished demand and price and complications in seafood transportation.

By the end of 2021, two of the five companies managed to return to their 2019 cluster. Some Covid-19 rules had been lifted and the country has embraced the new normal since October 2021. Conversely, three of the downgraded firms did not return to their original cluster. This suggests that the pandemic effects have proved to be long lasting for some firms.

6. Conclusion

The outbreak of Covid-19 has created a unique setting of considerable uncertainty in investigating firm performance, which motivates the writing of this article. Clustering was performed with the CoDa methodology on two key industries in Vietnam during the Covid-19 pandemic. Drawing from the results of this article, managers can compare the performance of their own firms with the profile of their closest cluster and possibly focus on management actions aimed at moving to a better performing cluster.

From a methodological point of view, the compositional approach prevents clusters from being very small (Yoshino *et al.*, 2016; Linares-Mustarós *et al.*, 2018) or being made up mainly of outliers, while introducing weights evens out the contributions of the financial statement figures (Jofre-Campuzano and Coenders, 2022). We repeated the classification in the standard way, in other words, by applying the Ward algorithm directly to the standard ratios in Table 1, and of the three obtained clusters one contained only three firms with outlying leverage ratios, and one contained only seven firms with outlying liquidity ratios.

Our hypotheses are all confirmed. First, three different company profiles were identified based on eight computed financial ratios that characterize different aspects of a firm's operations (solvency and profitability). Our study used data from two related industries with similar financial structures for which a single classification could be applied. As expected, the best-performing firms are those with lower levels of indebtedness and higher profitability (Cluster 1). In contrast, the worst cluster contains firms with the poorest performances regarding ROA, ROE, turnover, and margin (Cluster 3). This means that the first and fourth hypotheses are empirically supported.

As reflected in our results, five firms transitioned to a worse performing cluster in 2020 (8% of the sample size). By 2021, firms in the two industries were on track for recovery, indicated by the transition of two firms (3% of the sample size) to a better cluster in 2021. However, three firms stayed in the worse cluster for at least one more year. This is in line with Apedo-Amah *et al.* (2020)'s findings, which indicated that the destructive effects of the pandemic on economic growth prospects in general and firms in particular may be visible not only in the short term but also in the long run. This implies that gradual transitions may take place within a longer time frame. Full recovery from the pandemic in Vietnam did not occur until 2022, for which data were unavailable at the time of writing this manuscript and when more transitions to the best-performing clusters would be expected to occur. This, to some extent, supports our second and third hypotheses.

In this article, only six account categories (Non - current assets, Current assets, Non - current liabilities, Current liabilities, Revenues, Costs) have been considered. This was deemed sufficient to study the major concepts of solvency and profitability. A finer subdivision could separately consider buildings and equipment within non-current assets, accounts receivable and stocks within current assets, financial costs within costs, and define additional meaningful ratios. The subdivisions cannot be too fine-grained to avoid small categories with too large a proportion of zeros to be imputed.

Several limitations of this research can be highlighted. First, further years could have been added to the analysis to study long-term trajectories. However, this would make the interpretation of the transition matrices more demanding. Second, the financial standards and the pandemic effects are industry-specific. So, the results in this article cannot be extended to other industries behaving differently from FB and fishery without further independent research. It is not advised to combine very heterogeneous industries in the same cluster analysis because cluster differences can be driven by industry specificities rather than performance. Finally, other accounting figures and financial statements, such as the statement of cash flows (Arimany-Serrat *et al.*, 2022)), could be included in a compositional analysis.

Our future research will focus on using mixed data, which considers both quantitative and qualitative variables, for clustering. Integrating qualitative and quantitative aspects is expected to provide a more detailed specification of company profiles (Saleh *et al.*, 2023). Another further research avenue is to use fuzzy clustering methods in order to identify hybrid financial performance and distress profiles (Molas-Colomer *et al.,* 2024). Ideally, a larger sample size and a longer time frame are recommended to examine the consistency and effect of time lag on firm transitions over the years and to investigate the possible long-term effects of Covid-19 on their performance. Separate analyses by a wider range of industries could be beneficial to assessing the severity of the pandemic's impact and the adaptability of local businesses.

References

Aitchison, J. (1986), *The Statistical Analysis of Compositional Data*, Chapman and Hall, London.

Apedo-Amah, M.C., Avdiu, B., Cirera, X., Cruz, M., Davies, E., Grover, A., Iacovone, L., *et al.* (2020), *Unmasking the Impact of Covid-19 on Businesses: Firm Level Evidence from across the World*, The World Bank.

Arimany-Serrat, N., Farreras Noguer, M.À. and Coenders, G. (2022), "New developments in financial statement analysis: liquidity in the winery sector", *Accounting*, Vol. 8 No. 3, pp. 355–366.

Arimany-Serrat, N., Farreras-Noguer, M.À. and Coenders, G. (2023), "Financial resilience of Spanish wineries during the Covid-19 lockdown", *International Journal of Wine Business Research*, Vol. 35 No. 2, pp. 346–364.

Boubaker, S., Liu, Z. and Zhan, Y. (2022), "Customer relationships, corporate social responsibility, and stock price reaction: Lessons from China during health crisis times", *Finance Research Letters*, Vol. 47, p. 102699.

Boubaker, S., Goodell, J. W., Kumar, S. and Sureka, R. (2023), "COVID-19 and finance scholarship: A systematic and bibliometric analysis", *International Review of Financial Analysis*, Vol. 85, p. 102458.

Brancati, E. and Brancati, R. (2020), "Heterogeneous shocks in the COVID-19 pandemic: Panel evidence from Italian firms", Available at SSRN 3597650.

Capece, G., Cricelli, L., Di Pillo, F. and Levialdi, N. (2010), "A cluster analysis study based on profitability and financial indicators in the Italian gas retail market", *Energy Policy*, Vol. 38 No. 7, pp. 3394–3402.

Carreras-Simó, M. and Coenders, G. (2021), "The relationship between asset and capital structure: a compositional approach with panel vector autoregressive models", *Quantitative Finance and Economics*, Vol. 5 No. 4, pp. 571–590.

Caruso, G., Gattone, S.A., Fortuna, F. and Di Battista, T. (2018), "Cluster analysis as a decision-making tool: a methodological review", in Bucciarelli, E., Chen, S-H. and Corchado, J.M. (Eds.), *Decision Economics: In the Tradition of Herbert A. Simon's Heritage: Distributed Computing and Artificial Intelligence, 14th International Conference*, Springer, Cham, pp. 48–55.

Caruso, G., Gattone, S.A., Fortuna, F. and Di Battista, T. (2021), "Cluster analysis for mixed data: An application to credit risk evaluation", *Socio-Economic Planning Sciences*, Vol. 73, p. 100850.

Chue, T. K. and Xu, J. K. (2022), "Profitability, asset investment, and aggregate stock returns", *Journal of Banking & Finance*, Vol. 143, p. 106597

Cirera, X., Cruz, M., Grover, A., Iacovone, L., Medvedev, D., Pereira-Lopez, M. and Reyes, S. (2021), *Firm Recovery during Covid-19*, World Bank, Washington, DC.

Coenders, G., Egozcue, J.J., Fačevicová, K., Navarro-López, C., Palarea-Albaladejo, J. and Pawlowsky-Glahn, V. (2023), "40 years after Aitchison's article 'The statistical analysis of compositional data'. Where we are and where we are heading", *SORT-Statistics and Operations Research Transactions*, Vol. 47 No. 2, pp. 207–228.

Cowen, S.S. and Hoffer, J.A. (1982), "Usefulness of financial ratios in a single industry", *Journal of Business Research*, Vol. 10 No. 1, pp. 103–118.

Creixans-Tenas, J., Coenders, G. and Arimany-Serrat, N. (2019), "Corporate social responsibility and financial profile of Spanish private hospitals", *Heliyon*, Vol. 5 No. 10.

Dahiyat, A. A., Weshah, S. R. and Aldahiyat, M. (2021), "Liquidity and solvency management and its impact on financial performance: Empirical evidence from Jordan", *The Journal of Asian Finance, Economics and Business*, Vol. 8 No. 5, pp. 135–141.

Dolnicar, S., Grün, B. and Leisch, F. (2018), *Market Segmentation Analysis: Understanding It, Doing It, and Making It Useful*, Springer Nature, Singapore.

Egozcue, J.J. and Pawlowsky-Glahn, V. (2016), "Changing the reference measure in the simplex and its weighting effects", *Austrian Journal of Statistics*, Vol. 45, pp. 25–44.

Egozcue, J.J. and Pawlowsky-Glahn, V. (2019), "Compositional data: the sample space and its structure", *TEST*, Vol. 28 No. 3, pp. 599–638.

Faello, J. (2015), "Understanding the limitations of financial ratios", *Academy of Accounting and Financial Studies Journal*, Vol. 19 No. 3, pp. 75–85.

Fernández-Cerezo, A., González, B., Izquierdo Peinado, M. and Moral-Benito, E. (2023), "Firm-level heterogeneity in the impact of the COVID-19 pandemic", *Applied Economics*, Vol. 55 No. 42, pp. 4946–4974.

Goldstein, H. and Healy, M.J.R. (1995), "The graphical presentation of a collection of means", *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, Vol. 158 No. 1, pp. 175–177.

Greenacre, M. (2018), *Compositional Data Analysis in Practice*, Chapman and Hall/CRC, New York.

Greenacre, M. and Lewi, P. (2009), "Distributional equivalence and subcompositional coherence in the analysis of compositional data, contingency tables and ratio-scale measurements", *Journal of Classification*, Vol. 26, pp. 29–54.

Halling, M., Yu, J. and Zechner, J. (2020), "How did Covid-19 affect firms' access to public capital markets?", *The Review of Corporate Finance Studies*, Vol. 9 No. 3, pp. 501–533.

Ho, K.-C., Huang, H.-Y., Pan, Z. and Gu, Y. (2023), "Modern pandemic crises and default risk: Worldwide evidence", *Journal of International Financial Management & Accounting*, Vol. 34 No. 2, pp. 211–242.

Hron, K., Engle, M., Filzmoser, P. and Fišerová, E. (2021), "Weighted symmetric pivot coordinates for compositional data with geochemical applications", *Mathematical Geosciences*, Vol. 53, pp. 655–674.

Hron, K., Menafoglio, A., Palarea-Albaladejo, J., Filzmoser, P., Talská, R. and Egozcue, J.J. (2022), "Weighting of parts in compositional data analysis: Advances and applications", *Mathematical Geosciences*, Vol. 54 No. 1, pp. 71–93.

Houmes, R., Jun, C. C., Capriotti, K., & Wang, D. (2018). Evaluating the long-term valuation effect of efficient asset utilization and profit margin on stock returns: Additional evidence from the DuPont identity. *Meditari Accountancy Research*, 26(1), 193-210.

Jofre-Campuzano, P. and Coenders, G. (2022), "Compositional classification of financial statement profiles: The weighted case", *Journal of Risk and Financial Management*, Vol. 15 No. 12, p. 546.

Jordà, Ò., Singh, S.R. and Taylor, A.M. (2022), "Longer-run economic consequences of pandemics", *Review of Economics and Statistics*, Vol. 104 No. 1, pp. 166–175.

Kao, C., Wang, Y.-Y., Ho, T.-C., Chen, Y.-S. and Chen, P.-C. (2023), "The impact of Covid-19 on the productivity of large companies in Taiwan", *Asia Pacific Management Review*, Vol. 28 No. 4, pp. 501–509.

Kong, T., Yang, X., Wang, R., Cheng, Z., Ren, C., Liu, S., Li, Z., *et al.* (2022), "One year after COVID: the challenges and outlook of Chinese micro-and-small enterprises", *China Economic Journal*, Vol. 15 No. 1, pp. 1–28.

Krylov, S. (2018), "Target financial forecasting as an instrument to improve company financial health", *Cogent Business & Management*, Vol. 5 No. 1, p. 1540074.

Landau, S., Leese, M., Stahl, D. and Everitt, B.S. (2011), *Cluster Analysis*, John Wiley & Sons, Chichester.

Lev, B. and Sunder, S. (1979), "Methodological issues in the use of financial ratios", *Journal of Accounting and Economics*, Vol. 1 No. 3, pp. 187–210.

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

Linares-Mustarós, S., Farreras-Noguer, M.À., Arimany-Serrat, N. and Coenders, G. (2022), "New financial ratios based on the compositional data methodology", *Axioms*, Vol. 11 No. 12, p. 694.

Liu, Z., Dai, P. F., Huynh, T. L., Zhang, T. and Zhang, G. (2023) "Industries' heterogeneous reactions during the COVID-19 outbreak: Evidence from Chinese stock markets", *Journal of International Financial Management & Accounting*, Vol. 34 No 2, pp. 243-278

Martín-Fernández, J.A., Hron, K., Templ, M., Filzmoser, P. and Palarea-Albaladejo, J. (2012), "Model-based replacement of rounded zeros in compositional data: classical and robust approaches", *Computational Statistics & Data Analysis*, Vol. 56 No. 9, pp. 2688–2704.

Martín-Fernández, J.A., Palarea-Albaladejo, J. and Olea, R.A. (2011), "Dealing with zeros", in Pawlowsky-Glahn, V. and Buccianti, A. (Eds.), *Compositional Data Analysis: Theory and Applications*, John Wiley & Sons, Chichester, pp. 43–58.

MFAT. (2022), "Viet Nam Food and Beverage Sector: An Overview of the Vietnamese Food and Beverage Sector", *New Zealand Embassy in Hanoi, Vietnam*.

Molas-Colomer, X., Linares-Mustarós, S., Farreras-Noguer, M.À. and 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*. Ahead-of-print.

Nguyen, T.D. (2022), *Food and Beverage: War to Drag down Post-Covid Recovery in 2022*.

Nguyen, T.L., Nguyen, H.T., Nguyen, K.C., Nguyen, T.P., Ha, T.H.G., Dao, T.G. and Vu, K.A. (2021), "Vietnam's food and beverage industry: 10 months of 2021", *Houselink Joint Stock Company*.

Noviyana, S., Febriyola, A.S. and Koranti, K. (2023), "Analysis of financial performance using liquidity ratio, solvency ratio, activity ratio, profitability ratio in pharmaceutical sub-sector manufacturing companies on the indonesia stock exchange for the 2018-2020 period in dki Jakarta", *Jurnal Syntax Transformation*, Vol. 4 No. 8, pp. 79–89.

Palarea-Albaladejo, J. and Martín-Fernández, J.A. (2008), "A modified EM alr-algorithm for replacing rounded zeros in compositional data sets", *Computers & Geosciences*, Vol. 34

No. 8, pp. 902–917.

Palarea-Albaladejo, J. and Martín-Fernández, J.A. (2015), “zCompositions—R package for multivariate imputation of left-censored data under a compositional approach”, *Chemometrics and Intelligent Laboratory Systems*, Vol. 143, pp. 85–96.

Parkes, G., Hart, M., Rudd, J. and Liu, R. (2018), “The role of behavioural competences in predicting entrepreneurial funding resource orchestration”, *Cogent Business & Management*, Vol. 5 No. 1, p. 1512833.

Pawlowsky-Glahn, V., Egozcue, J.J. and Tolosana-Delgado, R. (2015), *Modeling and Analysis of Compositional Data*, John Wiley & Sons, Chichester.

RL, M. and Mishra, A.K. (2022), “Measuring financial performance of Indian manufacturing firms: application of decision tree algorithms”, *Measuring Business Excellence*, Vol. 26 No. 3, pp. 288–307.

Rodrigues, L. and Rodrigues, L. (2018), “Economic-financial performance of the Brazilian sugarcane energy industry: An empirical evaluation using financial ratio, cluster and discriminant analysis”, *Biomass and Bioenergy*, Vol. 108, pp. 289–296.

Sáenz, J.V., Quiroga, F.M. and Bariviera, A.F. (2023), “Data vs. information: Using clustering techniques to enhance stock returns forecasting”, *International Review of Financial Analysis*, Vol. 88, p. 102657.

Saleh, I., Marei, Y., Ayoush, M. and Afifa, M.M.A. (2023), “Big Data analytics and financial reporting quality: qualitative evidence from Canada”, *Journal of Financial Reporting and Accounting*, Vol. 21 No. 1, pp. 83–104.

Saus-Sala, E., Farreras-Noguer, À., Arimany-Serrat, N. and Coenders, G. (2021), “Compositional DuPont analysis. A visual tool for strategic financial performance assessment”, in Filzmoser, P., Hron, K., Martín-Fernández, J.A. and Palarea-Albaladejo, J. (Eds.), *Advances in Compositional Data Analysis: Festschrift in Honour of Vera Pawlowsky-Glahn*, Springer, Cham, pp. 189–206.

SEAFDEC. (2022), *Fisheries Country Profile: Viet Nam*, Southeast Asia Fisheries Development Center.

Sharif, A., Aloui, C. and Yarovaya, L. (2020), “Covid-19 pandemic, oil prices, stock market, geopolitical risk and policy uncertainty nexus in the US economy: Fresh evidence from the wavelet-based approach”, *International Review of Financial Analysis*, Vol. 70, p. 101496.

Sharma, A., Adhikary, A. and Borah, S.B. (2020), “Covid-19′ s impact on supply chain decisions: Strategic insights from NASDAQ 100 firms using Twitter data”, *Journal of Business Research*, Vol. 117, pp. 443–449.

Soliman, M.T. (2008), “The use of DuPont analysis by market participants”, *The Accounting Review*, Vol. 83 No. 3, pp. 823–853.

Stevens, S.S. (1946), “On the theory of scales of measurement”, *Science*, Vol. 103, pp. 677–680.

To, V.P., Nguyen, Q.K., Nguyen, L.T. and Vu, N.K. (2023), “Impacts of the Covid-19 pandemic on Vietnam’s marine fisheries”, *Asian Fisheries Science*, Vol. 36, pp. 68–78.

Toly, A.A., Permatasari, R. and Wiranata, E. (2020), “The effect of financial ratio (Altman z-score) on financial distress prediction in manufacturing sector in Indonesia 2016-2018”, *Advances in Economics, Business and Management Research*, Vol. 144, pp. 47–53.

VASEP. (2022), *Fishery Profile*, Vietnam Association of Seafood Exporters and Producers.

VIRAC. (2021), "How Vietnam Beverage industry react to pandemic in 2021", *Vietnam Industry Research and Consultancy*.

Ward Jr, J.H. (1963), "Hierarchical grouping to optimize an objective function", *Journal of the American Statistical Association*, Vol. 58 No. 301, pp. 236–244.

WHO. (2023), *Viet Nam COVID-19 Situation Report, Number 107*.

Williams, G.J. and Huang, Z. (1997), "Mining the knowledge mine: The hot spots methodology for mining large real world databases", in Sattar, A. (Ed.), *Advanced Topics in Artificial Intelligence: 10th Australian Joint Conference on Artificial Intelligence, AI'97 Perth, Australia, November 30--December 4, 1997 Proceedings 10*, Springer, Berlin, pp. 340–348.

Yeo, A.C., Smith, K.A., Willis, R.J. and Brooks, M. (2001), "Clustering technique for risk classification and prediction of claim costs in the automobile insurance industry", *Intelligent Systems in Accounting, Finance & Management*, Vol. 10 No. 1, pp. 39–50.

Yoshino, N., Taghizadeh-Hesary, F., Charoensivakorn, P. and Niraula, B. (2016), "Small and medium-sized enterprise (SME) credit risk analysis using bank lending data: An analysis of Thai SMEs", *Journal of Comparative Asian Development*, Vol. 15 No. 3, pp. 383–406.

Appendix 1: Compositional data analysis (CoDa)

A compositional vector and the simplex

A compositional vector is a vector with D strictly positive parts in which the only relevant information is contained in the ratios between said parts. Aitchison (1986) has developed the *Aitchison geometry* concept to work with compositional vectors. This geometry requires specific definitions and a specific metric to become Euclidean, which is essential for cluster analysis. Consider the compositions $\boldsymbol{x} = (x_1, x_2,..., x_D)$, $\boldsymbol{y} = (y_1, y_2,..., y_D)$ and a real number $\alpha > 0$.

- Perturbation of $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined as

$$\boldsymbol{x} \oplus \boldsymbol{y} = (x_1 y_1, ..., x_D y_D).$$

- Powering of $\boldsymbol{x}$ and α as:

$$\alpha \odot \boldsymbol{x} = (x_1^{\alpha}, ..., x_D^{\alpha}).$$

- Scalar product of $\boldsymbol{x}$ and $\boldsymbol{y}$ as:

$$< \boldsymbol{x}, \boldsymbol{y} >_A = \frac{1}{2D}\sum_{i=1}^{D}\sum_{j=1}^{D}\log\frac{x_i}{x_j} \times \log\frac{y_i}{y_j} = \sum_{i=1}^{D}\log\frac{x_i}{g(x)} \times \log\frac{y_i}{g(y)},$$

where

$$g(x) = \sqrt[D]{x_1 \cdots x_D}$$

is the geometric mean of $\boldsymbol{x}$.

- The Aitchison compositional distance between $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined by

$$d_A(\boldsymbol{x},\boldsymbol{y}) = \left(\frac{1}{D}\sum_{i=1}^{D-1}\sum_{j=i+1}^{D}\left(\log\frac{x_i}{x_j} - \log\frac{y_i}{y_j}\right)^2\right)^{\frac{1}{2}} = \left(\sum_{i=1}^{D}\left(\log\frac{x_i}{g(x)} - \log\frac{y_i}{g(y)}\right)^2\right)^{\frac{1}{2}}.$$

The compositional space becomes a vector space with the perturbation and the powering operations.

Transformations into coordinates

Compositional vectors cannot be analyzed directly. These vectors have to be transformed into a coordinate space which preserves all the information in the composition. There are several possible representations (Pawlowsky-Glahn *et al.*, 2015) of which the following are the most useful for cluster analysis:

1. Centered logratios (CLR)

$$CLR(\boldsymbol{x}) = (CLR_1, CLR_2,..., CLR_D) = \left(\log\frac{x_1}{g(x)}, ..., \log\frac{x_D}{g(x)}\right).$$

2. Isometric logratios (ILR): The ILR do not have a unique algebraic expression. They can be expressed as scaled logs of ratios between the geometric means of non-overlapping groups of parts, called balances. Each possible way of grouping the parts yields a different set of ILR.

We refer the reader to (Egozcue and Pawlowsky-Glahn, 2019; Greenacre, 2018; Pawlowsky-Glahn *et al.*, 2015) for insightful discussions of the comparative advantages of each representation.

Properties of centered logratios (CLR)

In this study, we use CLR, which have the following properties. Consider $\boldsymbol{x}$, $\boldsymbol{y}$ and the real constants α, β. It holds that:

(a) $\mathrm{CLR}(\alpha \odot \boldsymbol{x} \oplus \beta \odot \boldsymbol{y}) = \alpha \cdot \mathrm{CLR}(\boldsymbol{x}) + \beta \cdot \mathrm{CLR}(\boldsymbol{y})$;

(b) $d_A(\boldsymbol{x},\boldsymbol{y}) = d_E(CLR(\boldsymbol{x}), CLR\ (\boldsymbol{y}))$.

Property (a) implies that the CLR is an isomorphism of vector spaces: operations in the simplex are translated into operations in a real vector space. Property (b), where d_E stands for the Euclidean distance, characterizes CLR as an isometry and all metric concepts in the simplex are maintained after a CLR-transformation. Due to its preserved distance, CLR can be applied in cluster analysis, for example Ward clustering, which is also based on the Euclidean distances in the dataset. The ILR representation also fulfils properties (a) and (b) and could be used in Ward clustering with identical results.

In this article we use the weighted variant of CLR and thus the weighted variant of Aitchison's distance (Greenacre, 2018).

Appendix 2: Boxplots

Insert Figure 2A about here

Insert Figure 3A about here

Table 1: Definition of the standard financial ratios

	Financial ratios	Formula
Ratios related to solvency	Liquidity ratio = $\frac{Current\ assets}{Current\ liabilities}$	$\frac{x_2}{x_4}$
	Debt-maturity ratio = $\frac{Long-term\ debt}{Total\ debts}$	$\frac{x_3}{x_3+x_4}$
	Indebtedness ratio = $\frac{Total\ debts}{Total\ assets}$	$\frac{x_3+x_4}{x_1+x_2}$
Ratios related to profitability	Turnover = $\frac{Revenues}{Total\ assets}$	$\frac{x_5}{x_1+x_2}$
	Margin = $\frac{Net\ income}{Revenues}$	$\frac{x_5-x_6}{x_5}$
	Return On Assets (ROA) =$\frac{Net\ income}{Total\ assets}$	$\frac{x_5-x_6}{x_1+x_2}$
	Leverage = $\frac{Total\ assets}{Equity}$	$\frac{x_1+x_2}{(x_1+x_2)-(x_3+x_4)}$
	Return On Equity (ROE) = $\frac{Net\ income}{Equity}$	$\frac{x_5-x_6}{(x_1+x_2)-(x_3+x_4)}$

Note: Total assets = Current assets + Non - current assets

Total debts = Non-current liabilities (Long-term debt) + Current liabilities

Equity = Total assets – Total debts

Net income = Revenues - Costs

Table 2: Data descriptive statistics

Year	**2019**	**2020**	**2021**
Number of observations	66	66	66
Input parts (billion VND)			
x_1 Non-current assets	2.56 (9.42)	2.79 (10.93)	2.58 (10.37)
x_2 Current assets	2.15 (4.90)	2.43 (5.75)	3.21 (7.90)
x_3 Non-current liabilities	0.37 (1.85)	0.95 (6.37)	0.93 (6.06)
x_4 Current liabilities	1.55 (4.30)	1.73 (5.24)	1.89 (5.06)
x_5 Revenues	4.34 (9.73)	4.91 (12.68)	5.16 (13.81)
x_6 Costs	3.80 (8.13)	4.43 (11.60)	4.57 (11.95)
Standard financial ratios (Table 1)			
1. Liquidity	1.90 (1.78)	2.04 (1.68)	1.98 (1.44)
2. Debt maturity	0.11 (0.16)	0.13 (0.18)	0.12 (0.18)
3. Indebtedness	0.45 (0.19)	0.45 (0.21)	0.46 (0.19)
4. Turnover	1.52 (1.05)	1.47 (1.17)	1.43 (1.32)
5. Margin	0.07 (0.08)	0.05 (0.10)	0.05 (0.10)
6. ROA	0.10 (0.13)	0.07 (0.08)	0.06 (0.07)
7. Leverage	2.15 (1.03)	2.32 (1.56)	2.39 (2.12)
8. ROE	0.17 (0.31)	0.09 (0.25)	0.10 (0.17)
Firm information			
Industry (%)			
Fishery (FI)	24.2	24.2	24.2
Food and Beverage (FB)	75.8	75.8	75.8
Firm age (years)	27.09 (13.51)	28.09 (13.51)	29.09 (13.51)
Ownership type (%)			
State-owned (SO)	39.4	39.4	39.4
Private-owned (PO)	60.6	60.6	60.6
Exchange listed (%)			
Hanoi Stock Exchange (HNX)	24.2	24.2	24.2
Ho Chi Minh Stock Exchange (HOSE)	34.8	34.8	34.8
Over the Counter (OTC)	1.5	1.5	1.5
Unlisted Public Company Market (UPCOM)	39.4	39.4	39.4
Year of listing	2010 (5.0)	2010 (5.0)	2010 (5.0)

Note: Means and standard deviation (in parentheses) are reported for input variables, financial ratios, firm age and year of listing.

Table 3: Arithmetic part means normalized to unit sum acting as weights

Part	**Mean**
Non-current assets	0.12
Current assets	0.16
Non-current liabilities	0.02
Current liabilities	0.10
Revenues	0.31
Costs	0.29

Note: Mean values are the weights for equation (5).

Table 4: Contributions of the parts to total CLR variance (%) labelled according to the accounting figure in the numerator

Part	Unweighted	Weighted
Non-current assets	0.03	0.15
Current assets	0.06	0.09
Non-current liabilities	0.75	0.58
Current liabilities	0.04	0.07
Revenues	0.06	0.06
Costs	0.06	0.06

Table 5: Cluster profiles: means of representative standard financial ratios

Financial ratios	Cluster 1 n = 26	Cluster 2 n = 55	Cluster 3 n = 117
Liquidity	2.22	2.06	1.33
Debt maturity	0.00	0.02	0.12
Indebtedness	0.37	0.38	0.43
Turnover	2.10	2.06	0.89
Margin	0.08	0.06	0.06
ROA	0.17	0.11	0.05
Leverage	1.59	1.61	1.74
ROE	0.27	0.18	0.09

Note: n is the number of cases in each cluster;

Table 6: Transition matrix for 2019-2020

Year 2019		Year 2020		
Cluster	Total	Cluster 1	Cluster 2	Cluster 3
1	10	8	1	1
2	19	0	16	3
3	37	0	0	37
Total		8	17	41

Note: The reported totals are the number of companies in each cluster. The diagonal of the matrix shows stable companies between 2019 and 2020.

Table 7: Transition matrix for 2020-2021

Year 2020		Year 2021		
Cluster	Total	Cluster 1	Cluster 2	Cluster 3
1	8	8	0	0
2	17	0	17	0
3	41	0	2	39
Total		8	19	39

Note: The reported totals are the number of companies in each cluster. The diagonal of the matrix shows stable companies between 2020 and 2021.

Figure 1: Steps in weighted compositional cluster analysis of financial statements

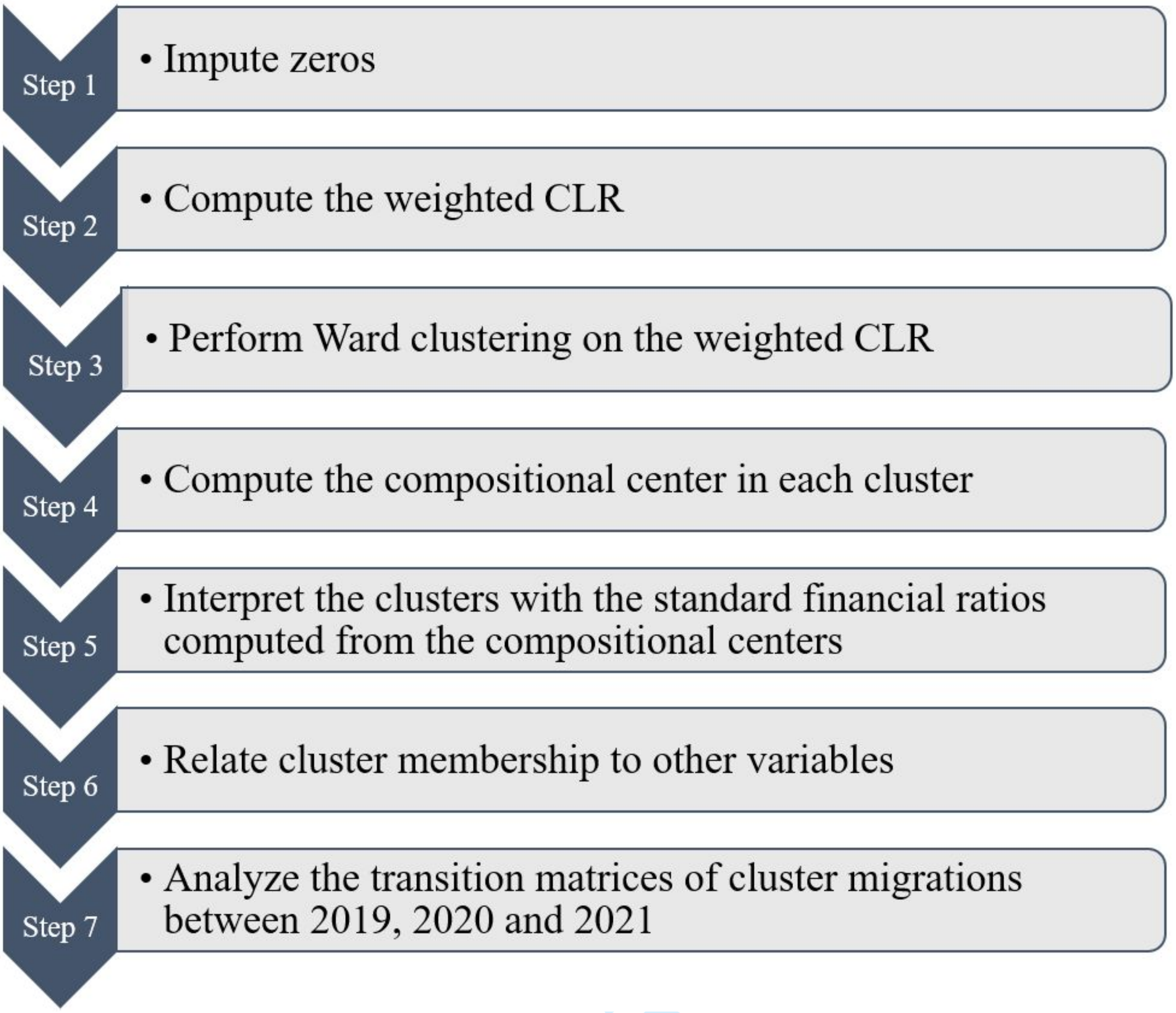


Figure 2A: Boxplots of the standard financial ratios

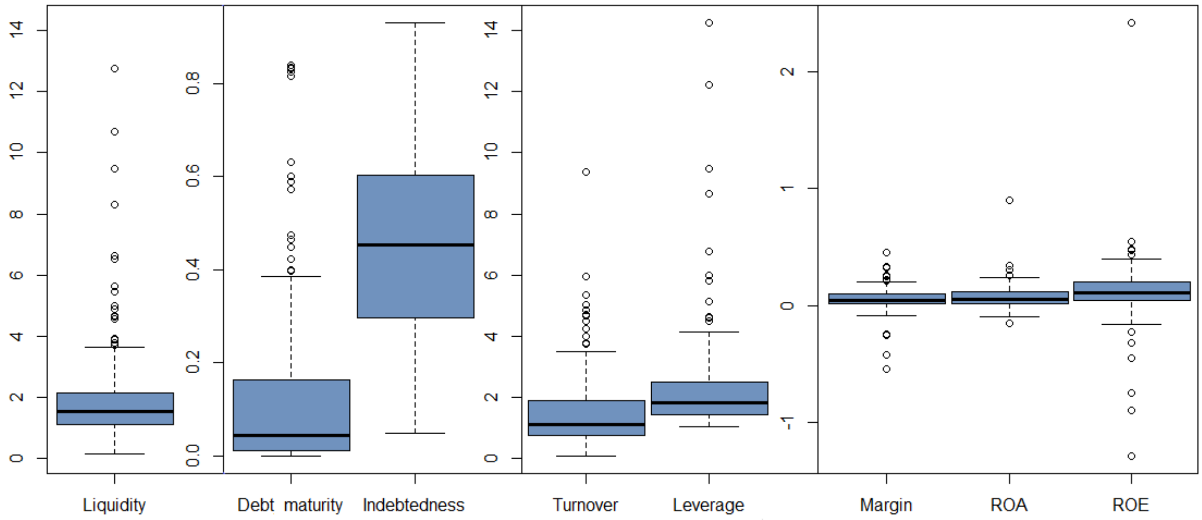


Figure 3A: Boxplots of financial statements transformed as weighted CLR labelled according to the accounting figure in the numerator

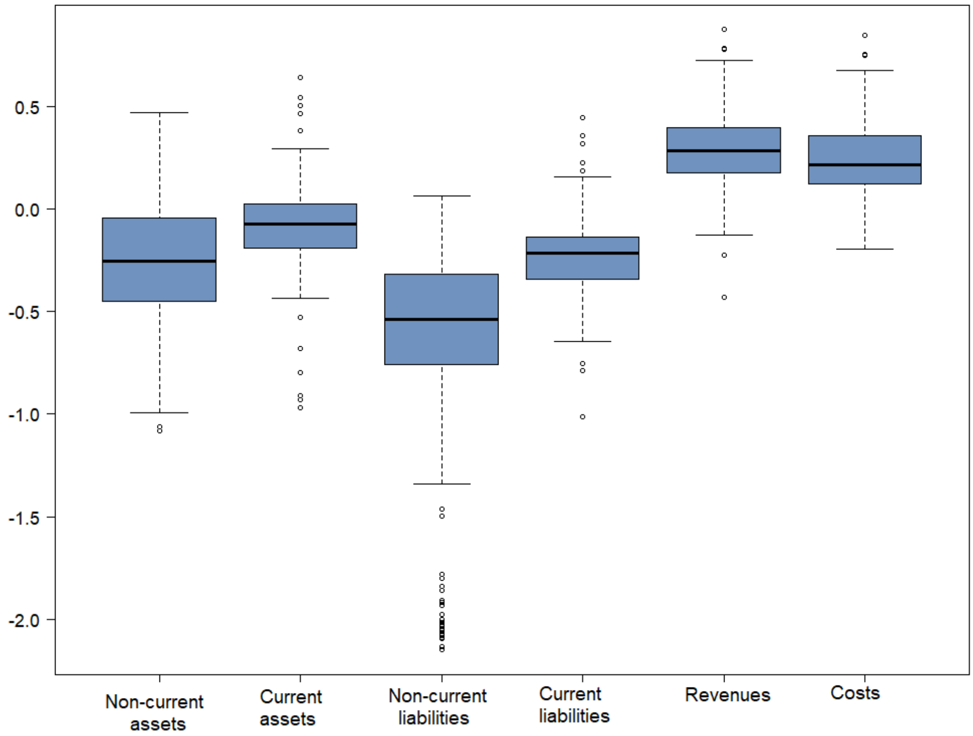


Figure 4: The weighted CLR classification: Ward dendrogram (Left) and scree plot of the within-cluster sums of squares (Right)

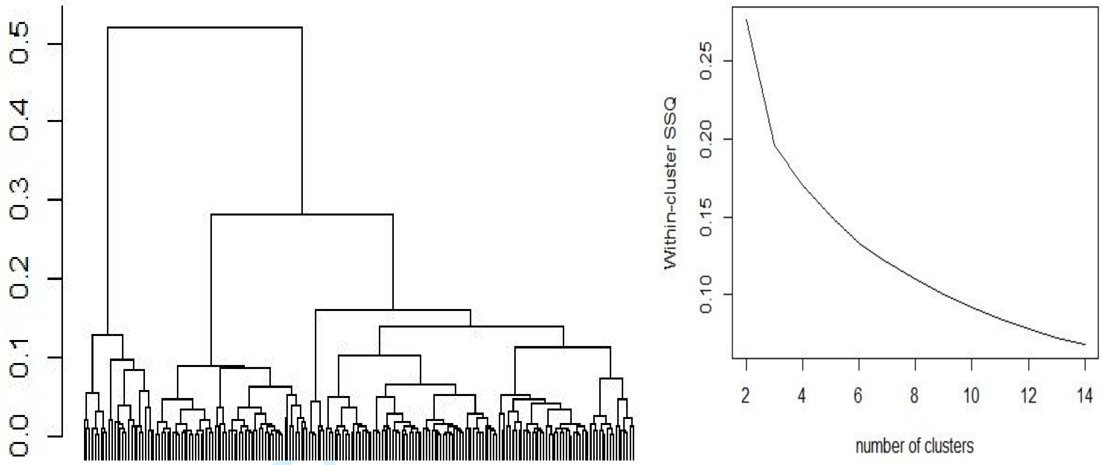
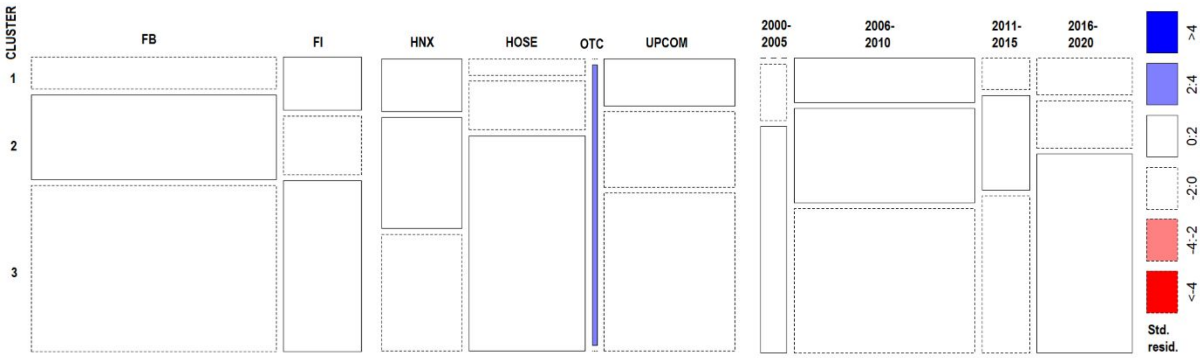


Figure 5: Mosaic plots of cluster membership and industry (left), type of stock exchange (center), and 5-year interval listing (right). Bar heights show percentages of cluster sizes within each category. Bar widths show category sizes. High standardized residuals show associations between cluster and category



Industry: FI (Fishery), FB (Food and Beverage); Exchange listed: HNX (Hanoi Stock Exchange), HOSE (Ho Chi Minh Stock Exchange), OTC (Over the counter), UPCOM (Unlisted public company market).

Figure 6: Confidence intervals in each cluster for the mean of log total assets (Left) and for the mean of log revenue (Right).