Anomaly Detection of Financial Ratios using Gaussian Mixture Models

Scrucca, L.
Department. of Economics (UNIPG)

Key-words: Gaussian mixtures, model-based clustering, unsupervised learning, entropy, outliers, financial ratios

JEL-codes: C13, C18, C38

Introduction

Financial stability is paramount for any organization's sustainability. Detecting anomalies in financial ratios can provide early warning signs of potential issues. This study leverages Gaussian mixture models by using an entropy-based anomaly detection technique (Scrucca, 2023), to enhance the accuracy and timeliness of anomaly detection. Inspired by Altman's seminal work on financial ratios and corporate bankruptcy prediction (Altman, 1968), this research aims to extend Altman's approach in an unsupervised manner in order to create a robust framework for anomaly detection. An application of the proposed methodology demonstrates promising results in detecting anomalies within financial ratios.

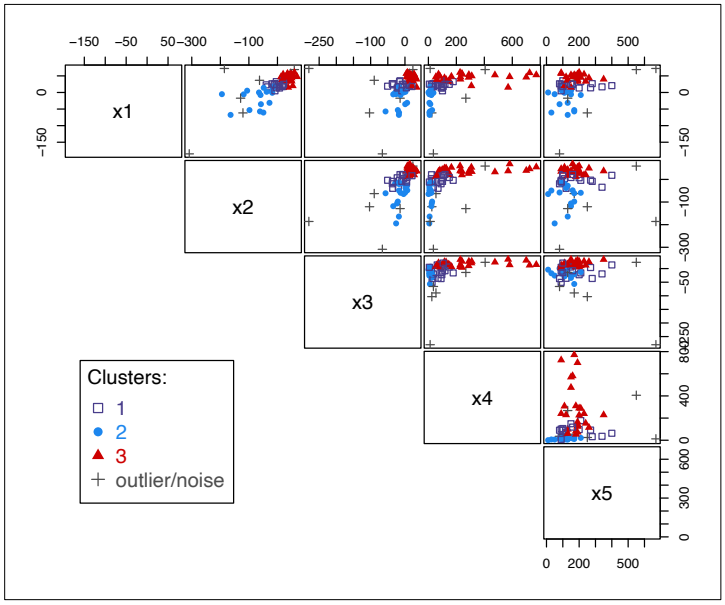
Methods

Gaussian mixture models (GMMs) are a versatile probabilistic modeling technique that can be used to simultaneously identify the underlying distribution of financial ratios and to assess the potential presence of outlying observations. GMMs are trained in an unsupervised manner, allowing to learn the inherent patterns in the data without relying on labeled examples. Furthermore, the entropy-based approach introduced by Scrucca (2023) can be integrated into the model for improved anomaly detection sensitivity and adaptability to various types of anomalies. The `mclust` and `mclustAddons` packages (Scrucca et al., 2023) for the R statistical environment (R Core Team, 2023) implement the proposed methodology and have been employed for the analysis discussed in this paper.

Results

The proposed approach is applied to a sample of 66 banks, 33 of them bankrupt and 33 solvent, using 5 ratios: working capital/total assets (x_1), retained earnings/total assets (x_2), earnings before interest and taxes/total assets (x_3), market value equity/book value of total liabilities (x_4), and sales/total assets (x_5). The scatterplot matrix shown in Figure 1 summarizes the final results, indicating the presence of three clusters and a small group of outliers. The table on the right-hand side cross-tabulates the estimated

clusters with the actual status of the banks (bankrupt vs solvent). Cluster 2 consists of all solvent banks, cluster 3 comprises all bankrupt banks, and cluster 1 is a mix of both types. Finally, the last cluster refers to the identified anomalies, which, as shown in the scatterplot matrix, are located at the outskirts of the data distribution.



	bankrupt	solvent
1	6	11
2	0	17
3	26	0
outlier/noise	1	5

Figure 1. Scatterplot matrix of financial ratios showing estimated clusters and identified outliers. Table on the right-hand side reports the cross-tabulation of identified groups with the reported status of the banks.

References

Altman, E. I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The Journal of Finance, 23:4, 589–609.

Scrucca, L., 2023. Entropy-based anomaly detection for Gaussian mixture modeling. Algorithms 16:4, 195.

Scrucca L., Fraley C., Murphy T. B. and Raftery A. E., 2023. Model-based clustering, classification, and density estimation using mclust in R. Chapman & Hall/CRC, New York.

R Core Team, 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

DOI 10.5281/zenodo.10053886