MICRO-PANEL CLUSTERING IN FINANCIAL DISTRESS PREDICTION

Mária Stachová – Pavol Kráľ

Abstract

We identify a company as being in financial distress if it cannot pay-off its financial obligations or it has difficulties to repay them. This situation is often considered as a dichotomous state, i.e. either a company is financially distressed, or it is not. However, we believe that this phenomenon is naturally multilevel one and the levels can be revealed by studying values of various financial indicators in time points preceding the occurrence of financial distress, focusing on changes in these indicators over time.

Based on our previous work, we decided to use: The Equity, Current ratio and Earnings after taxes as three financial ratios defining the financial distress of a company.

In our contribution, we analyze the past trends of these selected financial indicators and their impact on financial distress of companies. We identify the patterns in these trends by using the two-step clustering algorithm named CluMP and compare the results with the K-means based clustering method KML. Focusing on similar patterns in trends of financial indicators from previous time periods, we can refine the rough distress definition.

Key words: Financial distress, panel data, clustering

JEL Code: C38, G33

Introduction

When we attempt to create financial distress prediction models, the core piece of the puzzle we need to solve is to select an appropriate operational definition of financial distress for data at hand which on one hand reflects the general concept of financially distressed enterprises, i.e. enterprises which cannot pay-off their financial obligations or has significant difficulties to repay them, and on the other hand takes into account restrictions of available enterprise data and their socio-economic environment specifics, i.e. is very context dependent and data driven. We usually base such definitions on combination of various financial and economic indicators connected to enterprises themselves and their socio-economic environment as well as legal definitions of financial distress and bankruptcy. As many researches focused their research on the area of financial distress modelling, we can find multiple attempts to provide such definition in the literature. Among others, Baixauli et al. (2010) used four financial indicators of profitability and expert opinion, Li and Liu (2009) based their definition on negative economic results for two consecutive years and Alfaro-Cid et al. (2009) for three consecutive years, Stachová et al. (2015) identified as financial distress situation in which a company went bankrupt, has some ongoing liquidation or has some overdue obligations. In the paper, we adopt definition financial distress used in (Boďa and Úradníček, 2016) where authors consider company being financially distressed if all the following conditions are fulfilled simultaneously

a) its Equity was negative,
b) its Earnings after taxes (EAT) was negative,
c) its Current ratio attained a value lower than 1.

The problem with all above mentioned definitions is that they are in their nature static ones and describe a state of the company in one particular time point neglecting a path the enterprise underwent to reach that state. As we often characterize the financial distress of the company via multiple financial indicators, we can associate a multidimensional longitudinal pathway to each company and we assume that closer inspection of this pathway via specialized micro-panel clustering algorithms could bring valuable additional pieces of information to the definition of financial distress at hand in the form of patterns identification in pathways of companies. This additional information could result in further modification or even rejection of that definition.

The aim of our contribution is to investigate feasibility of such pathway inspection in the case of Slovak enterprises using two different clustering algorithms: micro-panel clustering algorithms: the two-step clustering algorithm CluMP and k-means based clustering algorithm KLM.

1 Data and Methodology

In our analysis, we used data that contains information of three numeric financial distress indicators – Earnings after taxes (EAT), Current ratio and Equity from 59 064 Slovak companies. These enterprises cover economic activities manufacturing, construction and wholesale according to SK NACE classification (C, F, G). The data are related to a range of five fiscal periods from 2013 to 2017.

We believe, that the significant changes in the values of each of these three (Equity, EAT and Current ratio value) criteria can indicate the changes in financial health of a monitored company especially if they decrease rapidly over the time.

In order to fulfill the goal of our paper we use two algorithms. First, the K-means partitioning to cluster the time trajectories of selected outcomes which is implemented in the R

(R Core Team, 2020) package “kml” (Genolini & al., 2015) using the function *kml*(). Second, the feature-based clustering algorithm implemented in the R package CluMP (Fojtík, Grishko and Sobíšek, 2020).

1.1 K-means clustering for longitudinal data

This subsection recollects a short introduction to K-means algorithm as presented in (Stachová and Sobíšek, 2016)

The K-means clustering applied in the “kml” package is modified for the longitudinal data. This algorithm is based on the original K-means clustering (MacQueen, 1967). This method minimizes the utility function iteratively for the time t, N objects according to an assumption of C clusters. The utility function can be expressed as follows:

$$\min \sum_{i=1}^{N} \sum_{c=1}^{C} u_{ict} d_{ict}^{2}, \tag{1}$$

where u_{ict} is a degree of appropriateness of the i-th object into the K-th cluster in the time t with conditions:

$$\sum_{c=1}^{C} u_{ict} = 1, \forall i, t, \tag{2}$$

$$\forall u_{ict} : u_{ict} = \begin{cases} 1 & \lVert \mathbf{x}_{it} - \mathbf{h}_{ct} \rVert = \arg\min_i \lVert \mathbf{x}_{it} - \mathbf{h}_{ct} \rVert \\ 0 & \text{elsewhere} \end{cases}.$$

We used the Euclidean distance $d_{ict} = \lVert \mathbf{x}_{it} - \mathbf{h}_{ct} \rVert$ between i-th vector of objects $\mathbf{x}_{it} = (x_{i1t}, \ldots, x_{ijt}, \ldots, x_{iJt})'$ and K-th centroid $\mathbf{h}_{ct} = (h_{c1t}, \ldots, h_{cjt}, \ldots, h_{cJt})'$ in the time t. We applied the algorithm to the standardized values of variables.

1.2 Feature-based partitioning for micro-panel data clustering (CluMP)

Feature-based clustering of micro-panel data (CluMP) algorithm is described in detail in working paper (Sobíšek, Stachová and Fojtík, 2018).

It is a two-step characteristic-based approach designed for this type of data. In the first step, the panel data are transformed into static data with lower dimension using a set of the proposed dynamic characteristics, representing different features of the time course of the observed variables. In the second step, the elements are clustered by clustering techniques designed for static data.

The used characteristics are:

- Average triangular difference between the two consecutive measurement values,

- Selective standard deviation of triangular differences between the two consecutive measurements,
- Average absolute triangular difference between the two consecutive measurements,
- Selective standard deviation of absolute triangular differences between the two consecutive measurements,
- Selective standard deviation of absolute triangular differences between the two consecutive measurements,
- Average growth coefficient,
- The ratio of positive to negative changes,
- The value of maximum angle between the line connecting peripheral measurements and the one between the inner point and the first measurement (in radians).

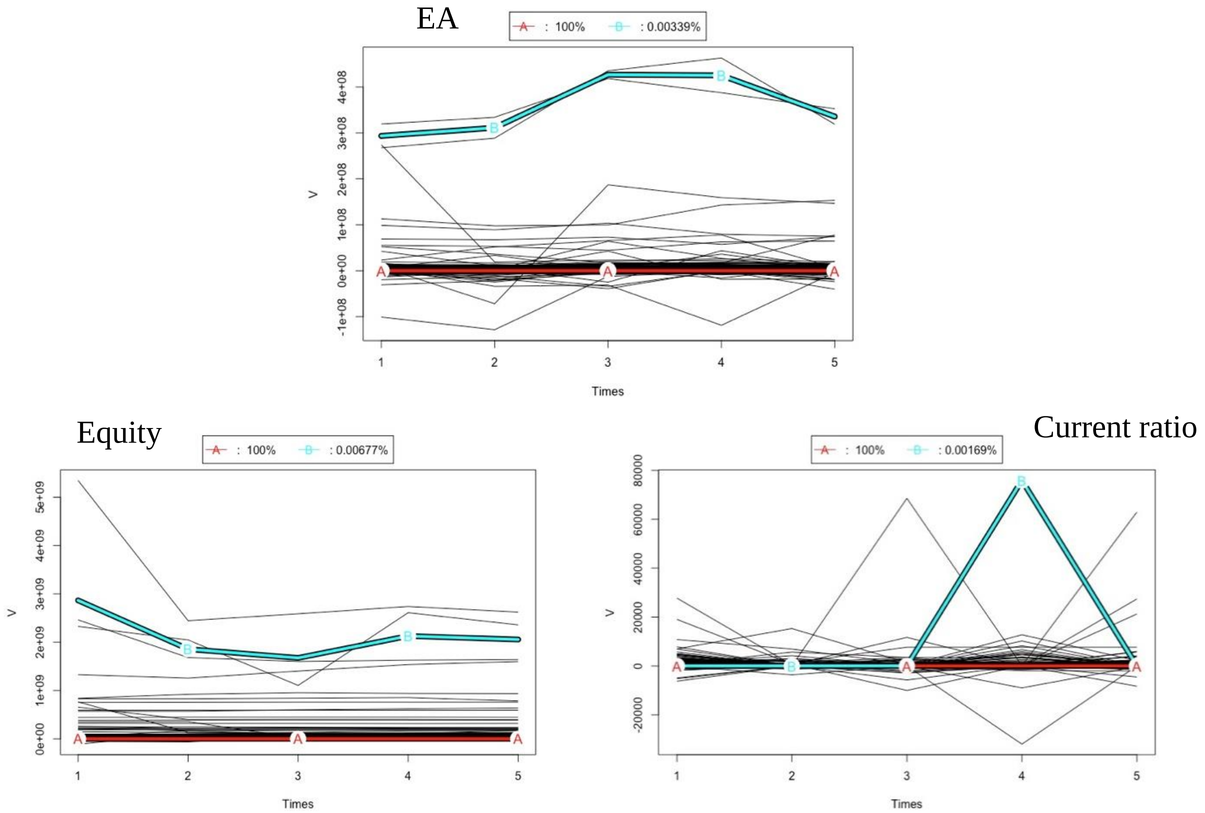
2 Results

If we follow the static expert's definition of financial distress as in (Boďa and Úradníček, 2016) and apply it to our data from the year 2017, we will find the 10 112 (17%) companies being in financial distress. In the next subsection we describe the results obtained by two mentioned clustering methods for longitudinal data.

We apply the K-means clustering on all three indicators and we obtained two clusters. The means of trajectories of created clusters and the percentage of values in each cluster are shown in Figure 1. Numbers 1-5 stand for years 2013-2017 respectively.

It can be seen that in case of all three indicators the "B" cluster contains slightly higher values. This fact could indicate that this cluster contains companies with lower risk of financial distress. But this could be misleading. As we can see on profile plots, this method acts very roughly as it created one large cluster and a very small second one. Moreover, the huge cluster "A" contains not only the companies that tend to be at the risk of financial distress, but also almost all financially healthy companies.

Fig. 1: Estimated cluster specific mean longitudinal profiles for EAT, Equity and Current ratio (K-means)

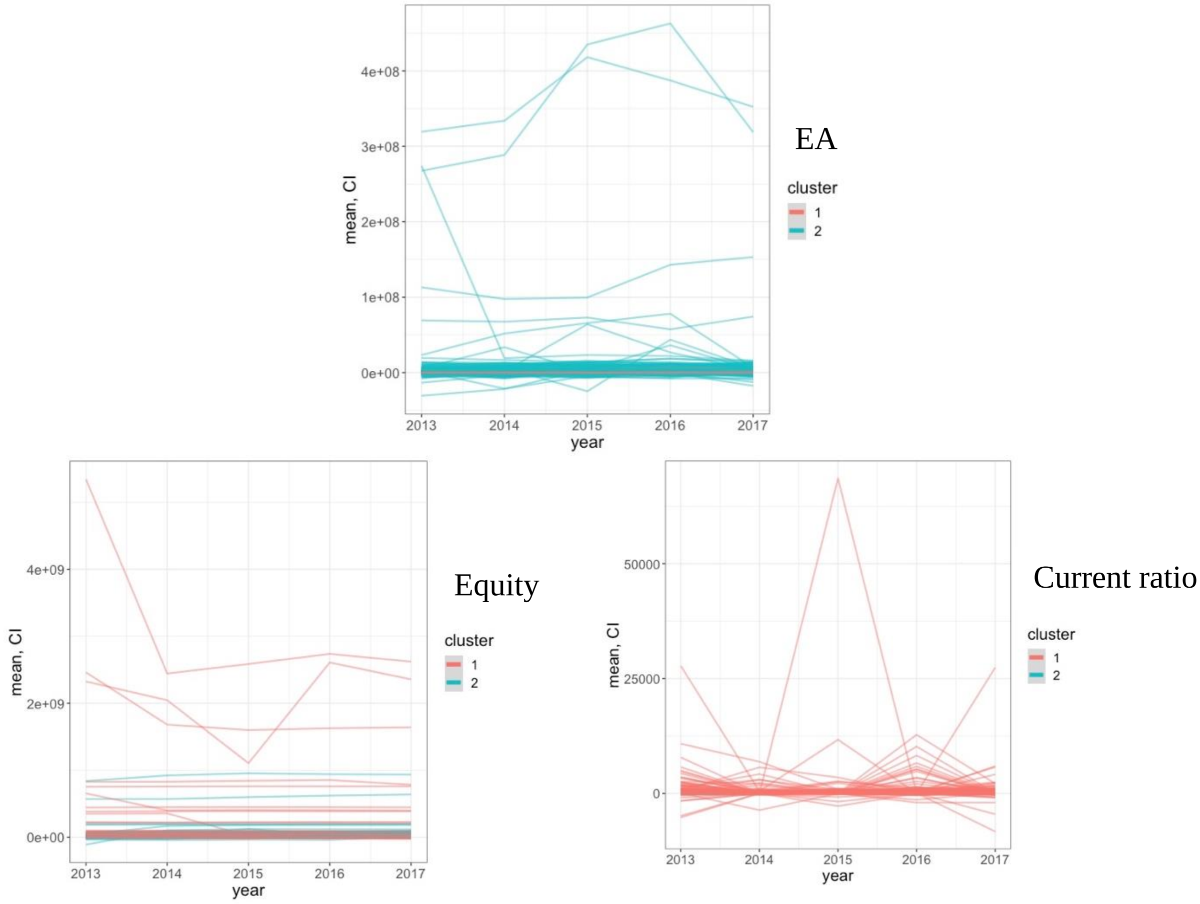


Source: The author's work

As the next step, we applied the CluMP clustering algorithm on all three indicators and we again obtained two clusters for each of them. The means of trajectories of created clusters and the percentage of values in each cluster are shown in Figure 2.

In case of EAT clustering, we obtained cluster "2" that contains 72% of companies and cluster "1" containing the rest. In the cluster "1", according the profiles mean, we can find also companies that are not labeled as being financially distressed from static's point of view, but they may end up in the state of distress because of their trend in time.

Fig. 2: Estimated cluster specific mean longitudinal profiles for EAT, Equity and Current ratio (CluMP)



Source: The author's work

In case of Equity clustering, we obtained cluster "1" that contains 70% of companies and cluster "2" containing the rest. The results are similar as those in case of EAT clustering.

In Figure 2, we can also see the results for Current ratio clustering. We obtained cluster "1" that contains 92% of companies and cluster "2" containing the rest. The partitioning is not as proportional as the results for EAT and Equity, but still more reasonable in comparison to results obtained by K-means clustering.

Conclusion

In the paper, we demonstrated identification of trend patterns in multidimensional longitudinal pathways formed by longitudinal trends of financial indicators of potentially financially distressed companies using two micro-panel clustering algorithms: the two-step clustering algorithm CluMP and k-means based clustering algorithm KLM. The presented results indicate the application of such algorithms can provide us with supportive information for further refinement of financial distress definition we used in the case of Slovak enterprises. They also seem to be in favor of using CluMP algorithm which was able to identify two meaningful clusters of pathways over KML method which failed to reveal any patterns of interest. Unfortunately, as only a one particular distress definition and restricted number of financial indicators were investigated, we cannot provide any definitive recommendation which of the algorithms utilized in the paper should be regarded as preferable. Therefore, our general recommendation is that multiple micro-panel clustering algorithms should be definitely used as a standard part of exploratory analysis when we attempt to predict financial distress to scrutinize or further refine the intended definition of financial distress. The results of such analysis can be seen as informal data driven validation that appropriate definition of financial distress has been chosen. In the future research, we would like to further investigate properties and potential of the micro-panel clustering algorithms using an extensive simulation study based on various plausible economic scenarios and supported by pieces of information extracted from available data of Slovak enterprises.

Acknowledgment

The work has been supported by project VEGA, No. 1/0767/18 SMART model – a decision support tool in management of enterprises.

References

Alfaro-Cid, E. et al. (2009). Finding Relevant Variables in a Financial Distress Prediction Problem Using Genetic Programming and Self-organizing Maps. In Computational Finance, 2, 31-49, ISNN 1860949X.

Baixauli, J. S., & Modica-Milo, A. (2010). The bias of unhealthy SMEs in bankruptcy prediction models. Journal of Small Business and Enterprise Development, 17(1), 60-77.

Boďa, M., & Úradníček, V. (2016). The portability of Altman's Z-score model to predicting corporate financial distress of Slovak companies. In Technological and Economic Development of Economy, 22(4), 532-553.

Fojtik, J., Grishko, A., & Sobisek, L. (2020). CluMP: Clustering of Micro Panel Data. Rpackage version 0.8. https://CRAN.R-project.org/package=CluMP

Genolini, Ch., Alacoque, X., Sentenac, M, & Arnaud, C. (2015). kml and kml3d: R Packages to Cluster Longitudinal Data. Journal of Statistical Software, vol. 65, iss. 4, pp. 1-34. URL http://www.jstatsoft.org/v65/i04/.

Li, D., & Liu, J. (2009). Determinants of Financial Distress of ST and PT Companies: A Panel Analysis of Chinese Listed Companies. Retrieved February 12, 2009, from http://ssrn.com/abstract=1341795

MacQueen, J., B. (1967). Some Methods for classification and Analysis of Multivariate Observations. Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, pp. 281-297.

Stachová, M. & Sobíšek, L. (2016). Financial distress criteria defined by clustering of longitudinal data. In: International Days of Statistics and Economics (MSED) [online]. Praha, 08.09.2016 – 10.09.2016, pp. 1703–1712. URL https://msed.vse.cz/msed_2016/article/248-Stachova-Maria-paper.pdf.

Sobisek, L., Stachova, M. & Fojtik, J. (2018). Novel Feature-Based Clustering of Micro-Panel Data (CluMP). arXiv e-prints arXiv:1807.05926.

R Core Team (2020): R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, http://www.r-project.org/

Contact

Mária Stachová

Faculty of Economics, Matej Bel University

Tajovského 10, 975 90, Banská Bystrica, Slovakia

maria.stachova@umb.sk

Pavol Kráľ

Faculty of Economics, Matej Bel University

Tajovského 10, 975 90, Banská Bystrica, Slovakia

pavol.kral@umb.sk