

## Expanded abstract

# The prediction of the business failure of the spanish cooperatives. Application of the Extreme Gradient Boosting Algorithm

Cooperatives are entities with regulatory and operational singularities that make them different from other commercial companies in the business landscape. The relations with the workers, the remuneration of the partner's contribution and the way in which the results are distributed configure them as companies of a deep social nature, even though they also have a commercial nature.

After reviewing the contents of the studies aimed at the identification of business failure prediction models, we have observed that there are still very few which address cooperative societies, despite the importance of this kind of organization in our business landscape. In addition, most of the papers follow traditional statistical methodologies. In fact, the different methodologies used in this field of research have evolved over time; and one of the newest and most innovative is *Extreme Gradient Boosting (XGBoost)*, which to the best of our knowledge has not been applied in this research area.

The objective of this study is to show the usefulness of Extreme Gradient Boosting methodology, which is based on machine learning techniques, in the prediction of business failure, particularly in the field of cooperative companies.

We will determine whether the algorithm is suitable for the identification of the most relevant variables that indicate situations of financial distress. Therefore, it will allow to classify a company as failed or not failed. We will also analyze whether this technique is appropriate to explain the reasons leading to this classification.

We also intend to contribute to overcoming the general absence in the literature of this type of studies in this business environment and to incorporate modern techniques based on artificial intelligence.

To achieve the aims of the study, a model has been estimated from a sample of non-financial cooperatives that failed in 2015, belonging to various productive sectors. To consider a company as failed, a broad definition of failure has been chosen, because it allows us to clearly distinguish between failed and non-failed companies. Thus, the failure has been identified to the legal definition of insolvency, as contemplated by the present Spanish insolvency regulation. The selection made, taking into account this condition, consisted in 56 companies that had filed insolvency proceedings in the year considered. Thereafter, they have been paired with as many non-failed ones with similar characteristics, so that the final sample has consisted of 112 companies.

Since our main goal is to formulate business failure prediction models, we have focused on those variables or financial ratios that provide information on aspects of the firm's solvency and profitability, but without forgetting the influence of the company's debts. To these main cate-

gories, we have added rotation, activity and asset structure. Given the uniqueness of this kind of companies we are considering, some clarifications have been made for some of the selected financial ratios, specifically for the debt and profitability ratios.

Results from the statistical study show the importance of the variables related to solvency, liquidity, debt and profitability. These variables exhibited the most contribution to detecting a situation of business failure.

The model combines in an effective way the variables of long-term solvency, liquidity, activity, and debt, and achieves a 100% accuracy rate in the training sample and 86% in an independent testing sample. The results have confirmed the higher the solvency and immediate liquidity of a cooperative company the lower the probability of failure. Additionally, the higher the variable that measures the ratio value added to sales the higher the probability of failure, as financial and personnel expenses are involved in this relationship. Finally, we have observed that profitability increases when cash flow grows in proportion to the company's assets, therefore being consistent in the sense that as a company performs better the propensity for failure is reduced by being more profitable.

These results are consistent with the empirical evidence: the higher the solvency and profitability rates the lower the failure probability, and the lower the debts the lower the exposure to failure. Consequently, we understand that the applied technique is quite effective for the objectives pursued, that is, the early detection of business failure for cooperative companies.

Probably the variable less easy to compare is profitability since in a cooperative the book result is not comparable with the rest of commercial companies because it does not take into account the potential benefits of their partners through prices of cooperative products and services.

The resulting model, in addition to classifying cooperative as failed and non-failed, has shown its potential to identify the underlying reasons for this classification. That is, the *XGBoost* model algorithm provides explanations about the variable's weights in terms of failure probability for each observation of the test sample, splitting the individual weight and the influence sign of each of the predictors. Therefore, the final model is not only about high accuracy predictions, but also leads to results that are fully interpretable—it is possible to determine the sign and the effect of each variable or financial ratio on the final prediction for each and every one of the observations.

The study is intended to be of a continuity nature because we consider it would be very interesting to carry out this type of analysis again but extending it to a series of several years in order to propose a model capable of detecting failure with a few years in advance.

In short, we have demonstrated that this study contributes to the literature on business failure in the cooperative sector by means of showing the utility of a new methodology (*XGBoost*) capable of identifying cooperatives with a high probability of failure. The results are very accurate, and their interpretation is very easy.

**Keywords:** Business failure, cooperatives, prediction models, Extreme Gradient Boosting.