



Multiperiod Bankruptcy Prediction Models with Interpretable Single Models

Ángel Beade¹ · Manuel Rodríguez^{1,2} · José Santos³ 

Accepted: 13 September 2023
© The Author(s) 2023

Abstract

This study considers multiperiod bankruptcy prediction models, an aspect scarcely considered in research despite its importance, since creditors must assess the risk of loans over the entire life of the debt and not at a specific point in the future. Two possibilities for the implementation of multiperiod prediction models are considered: Multi-Model multiperiod Bankruptcy Prediction Models (MMBPM) and Single-Model multiperiod Bankruptcy Prediction Models (SMBPM). The former considers the conditional probabilities obtained by individual models predicting bankruptcy at specific times in the future, while the latter is a single model predicting bankruptcy at a specific time interval in the future. The results show that there are no significant differences between the two approaches when compared using data after the learning period. However, SMBPMs have the important advantage of interpretability for decision-making, which is discussed with examples. Moreover, a comparison of SMBPM performance with external references is performed.

Keywords Business failure · Multiperiod · Explainable Artificial Intelligence · Interpretability · Genetic programming

1 Introduction

The need for multiperiod Bankruptcy Prediction Models (BPMs) (which make predictions for a future time interval rather than for a specific moment in time) arises quite naturally because it responds to the fact that creditors must face credit risk, not at a specific moment in time, but over the entire life of the debt. Despite this,

✉ José Santos
jose.santos@udc.es

¹ Business Department, University of A Coruña, Campus de Elviña, s/n, 15071 A Coruña, Spain

² Cátedra AECA-Abanca, UIE University, Santiago de Compostela, Spain

³ CITIC (Centre for Information and Communications Technology Research), Department of Computer Science and Information Technologies, University of A Coruña, Campus de Elviña, s/n, 15071 A Coruña, Spain

the use of multiperiod BPMs (especially in the long term) is rather limited and their research interest rather low (Aziz & Dar, 2006; Matenda et al., 2021).

There are two concepts to focus on (Luo et al., 2022): the prediction horizon and the task of a multiperiod BPM. At time t , the prediction horizon T refers to a future prospective time period. The multiperiod BPM estimates the probability of business failure occurring within the time period $[t+1, t+T]$. That is, multiperiod failure prediction consists of creating multiperiod BPMs that, e.g., with data from 2008 ($t=2008$), estimate the probability that a company will fail at different prediction horizons. For example, if the prediction horizon is 2 years, the interval will be [2009, 2010], if the prediction horizon is 5 years, the interval will be [2009, 2013]) and with a prediction horizon of 9 years, the multiperiod BPM will estimate the probability that a company will fail in the years from 2009 to 2017, inclusive.

The current approaches to address the realisation of multiperiod failure prediction (multiperiod BPM) share the same hypothesis: the probability of an event occurring (future) depends only on the current state (present) and not on previous states (past). For this reason, they use the calculation of the conditional probabilities of failure in a period without having failed in the previous period, probabilities which are taken from individual BPMs that predict failure at specific points in time in the future. In this way, the creation of a monotonically increasing time structure of the cumulative probability of failure is ensured, which makes the model plausible. Take, for example, the calculation of a multiperiod BPM with a prediction horizon of 9 years. To perform the failure prediction for the period $[t+1, t+9]$, the marginal probability of failure in years $t+1$; $t+2$; ...; $t+9$ has to be modelled. This implies, at least, a series of limitations that should be highlighted:

- The interpretability of the model is greatly compromised. Calculating the prediction of failure in the period $[t+1, t+9]$ on the basis of 9 marginal models assumes that these 9 models may have different explanatory variables, or, if they are similar, the relevance of each is likely to be different in each model. The prediction $[t+1, t+9]$ may perform well, but its interpretability will be far from desirable.
- Each of the 9 models implies a different output to which a probability function, different in each of them, needs to be fitted. Each of these adjustments involves a risk of bias in the prediction $[t+1, t+9]$.

In this context, our proposal is to perform discrete-time multiperiod BPMs without the need to model the marginal probability of failure in each of the periods of the prediction horizon, in which only one model directly predicts failure at a given time interval in the future. Returning to the previous example, in our proposal the multiperiod BPM with prediction horizon $[t+1, t+9]$ will not require models that assess the probability of failure at $t+1$; $t+2$; ...; $t+9$. Specifically, in our work, models with prediction horizons of 3, 6 and 9 years (intervals 1–3, 1–6 and 1–9 years prior to failure) will be considered. In addition, according to the above, models with the same prediction horizon obtained by the following two approaches will be compared:

- Those obtained by modelling the marginal probability of failure, which we will call Multi-Model multiperiod Bankruptcy Prediction Models or MMBPMs. In these models, to obtain the model with prediction horizon X (interval $[t+I, t+X]$), the independent BPMs that make predictions at a specific moment in time, (e.g., 1 year prior to failure or failure in exactly $t+I, \dots, X$ years prior to failure or failure in exactly $t+X$) will be used (i.e., X BPMs). With these independent BPMs, the probabilities of failure at each specific moment will be obtained and, from them, the probability of failure (cumulative probability of failure) at prediction horizon X (multiperiod failure prediction in the interval $[t+I, t+X]$).
- Those obtained directly without that modelling, which we will refer to as Single-Model multiperiod Bankruptcy Prediction Models or SMBPMs. Consequently, the time structure of the cumulative probability of failure over the prediction horizon is not used, providing only the cumulative probability of failure over that time interval, in exchange for ease of interpretability.

The Genetic Programming (GP) technique (Koza, 1992; Poli et al., 2008) is used for automatically learning all the prediction models in the different approaches. The chosen way to analyse the suitability of SMBPMs is to compare their performance with that of MMBPMs on the basis of data after the training period. Furthermore, a comparison of the performance of the SMBPMs during the training period (evaluated on the test set) with the results obtained in other studies will be carried out in order to assess the suitability of the technique used (GP).

The objective of this study is to obtain SMBPMs with two main characteristics:

- *Performance* High predictive power after the learning period, comparable to the corresponding MMBPMs.
- *Interpretability for decision-making* SMBPMs should allow (by applying the proposed and appropriate techniques) to explain and make sense of these models, as well as help to understand what actions need to be taken to counteract the failure of a company.

Therefore, the main objective and contribution of this study is to present an alternative to the use of annual models for obtaining multiperiod BPMs, easily interpretable for decision-making purposes and with long forecasting horizons (up to 9 years).

The rest of the article is structured as follows. Section 2 summarizes, in the area of business failure prediction, previous proposals on multiperiod BPMs. Section 3 details the design of the prediction models, the data used to derive the models, the set of explanatory variables, the genetic programming technique and the software environment. Section 4 focuses on the methods for obtaining MMBPMs and SMBPMs. Section 5 details the results of SMBPMs in terms of performance and interpretability, with a comparison between MMBPMs and SMBPMs, as well as including the comparison of SMBPMs with external references. Finally, Sect. 6 includes the main conclusions that can be drawn from this study.

2 Related Works

Early BPMs made predictions 1 year in advance of failure as shown in the work of Beaver (1966), Altman (1968) or Edmister (1972). This time horizon is still very frequent today (Kim & Upneja, 2021; Muslim & Dasril, 2021; Tsai et al., 2021). In any case, the need to expand the prediction horizon appears from the very beginning (Deakin, 1972; Wilcox, 1973) and this concern continues to the present day (Altman et al., 2020).

This evolution was reflected in the regulations and guidelines that concern financial institutions. Thus, while Basel II (International convergence of capital measure and capital standards: a revised framework) (Basel Committee on Banking Supervision, 2004) stipulated that an External Credit Assessment Institution (ECAI) should have been established for predicting failure with at least 1 year and preferably 3 years to be eligible for national supervisors, the guiding principles for the replacement of International Accounting Standards (IAS 39) (Basel Committee on Banking Supervision, 2009) recommended financial institutions to estimate the risk of a loan over the entire duration of the loan. Basel III (Basel Committee on Banking Supervision, 2011, 2017) does not change this criterion. On 27 October 2021, the European Commission adopted the “Banking Package” (informally known as Basel IV) which proposes an overhaul of EU banking rules (the Capital Requirements Regulation and the Capital Requirements Directive) from 2025 onwards. In any case, nothing suggests that the criterion that “financial institutions will estimate the risk of a loan over the entire duration of the loan” will undergo changes.

In the same sense, the entities that set the accounting rules also established that the measure of default risk should be based on the total life of the financial instrument (Financial Accounting Standard Board (FASB), 2016; International Accounting Standards Board, 2014).

Therefore, in parallel with the concern to extend the time horizon of prediction, the basis for research in multiperiod models arise. Following Blümke (2022), the literature on modelling multiperiod prediction of failure risk can be divided into three main possibilities:

- Survival models. Survival analysis is a set of methods commonly used to analyse data on the time elapsed until the occurrence of an event under study, such as the time until someone dies of a disease, is promoted in the job, etc. The focus is on modelling the prediction of the transition of the event and the time it takes for the event to occur.
 - In continuous time. With conditional probabilities with Poisson series (Duffie et al., 2007), or using the forward intensity approach (Duan & Fulop, 2013; Duan et al., 2012) or one of its modifications.
 - In discrete time. Sometimes a survival analysis is needed for discrete-time data, since, in practice, data is often collected at discrete time intervals, for example, days, weeks, and years, which violates the assumption of continuous time. On the other hand, the use of discrete time analysis has certain advantages, compared to its continuous-time counterpart. For example, discrete-

time analysis has no problems in dealing with multiple events occurring at the same point in time, nor does it present problems when the exact time of occurrence of the event is unknown (Chava & Jarrow, 2004; Shumway, 2001; Traczynski, 2017).

- Models based on transition matrices. The estimation of the probability of transition (change from one state to another) from one period to another is the central focus of these models. The works of Christensen et al. (2004), dos Reis and Smith (2018) and Jarrow et al. (1997) are representative of this possibility, all of which focus on the issue of bond rating, as a consequence of the credit risk and—ultimately—the problem of business failure.

It is observed that the common feature of the different approaches, both survival models and models based on transition matrices, is that the probability of an event occurring depends on the previous event (concept of conditional probability and Markov chains associated with transition matrices). Since, in our case, with multi-period BPMs we do not model the probability of state transitions but the prediction of an event in a time interval (similar to the survival model), our work could be framed within this typology of discrete-time survival models. Therefore, this work will use the studies by Duan et al. (2012), Duan and Fulop (2013), Luo et al. (2022) and Orth (2013) to evaluate our proposal (Sect. 5.1). All of them deal with multi-period survival models based on monthly data from US public companies and their approach is closer to the approach of this study than the one corresponding to the transition matrices.

3 Design of Prediction Models

Briefly, the main objective in this study is to obtain SMBPMs with prediction horizons of 3 years, 6 years and 9 years (which make predictions in the intervals [1, 3], [1, 6] and [1, 9] prior to failure), which have high performance in their predictions over time horizons and economic environments other than those of the learning period and are interpretable for decision-making. The main aspects involved in the design of these models are described in detail in the following subsections.

3.1 Data Sample

This study is based on the empirical analysis of the mortality of medium-sized Spanish firms. For the modelling, a population of 11,158 firms (1067 classified as failed and 10,091 classified as non-failed) with accounting information from 2005 to 2019 is available. The accounting information of the companies has been obtained from the SABI database (Iberian Balance Sheet Analysis System, <https://www.informa.es/en/business-risk/sabi>) of the company Informa, SA. It should be noted that the size of the population is larger than usual in other studies.

On the other hand, information from the Public Insolvency Register (www.publicidadconcurzal.es) and from companies specialised in business reports

has been used to obtain the specific legal information of their failure status. The legal information on failure status is available from 1 January 2005 to 31 December 2020.

The notion of medium-sized companies is understood to refer to the definition of micro, small and medium-sized companies, published in the Official Journal of the European Union L 124 of May 20, 2003. In the case of this work, the following criteria were applied jointly:

1. Workforce (measured in annual work units) equal to or greater than 50 and less than 250.
2. And at least one of the following conditions:
 - a. Annual turnover greater than EUR 10 million and less than or equal to EUR 50 million.
 - b. Annual balance sheet greater than EUR 10 million and less than or equal to EUR 43 million.

Due to the peculiarities of some sectors which, among other reasons, use specific accounting or valuation criteria, it has been decided to exclude them from the study so as not to alter the interpretation of the financial ratios (input information for the forecasting models) and distort the results. Efforts have been made to minimise the number of sectors to be excluded. The excluded sectors were as follows: building construction; civil engineering; specialised construction activities; financial services, including insurance, reinsurance and pension funds; activities auxiliary to financial services and insurance; finally, compulsory social security and general government activities. The study has also been limited to the following legal forms: public limited companies, private limited companies (limited liability companies) and co-operatives.

Observations of failed and non-failed companies from 2005 to 2007 (both included) are used to learn the different models (“Observation” means data referring to a firm in a financial year). Data from 2008 to 2019 are used to make the comparisons between SMBPMs and MMBPMs. This distribution of data is entirely intentional and allows to test the suitability of the models when used after the learning period and in economic environments clearly different from the period in which they were trained. This, together with the length of the comparison period, allows more reliable conclusions to be drawn.

In the context of this study, the concept of failure is related to the legal declaration of suspension of payments or bankruptcy. Given the multiple causes of failure, the legal declaration of suspension of payments or bankruptcy in research is, in the absence of confirmation of the event, the most widely used concept in studies of business failure due to its objectivity and concreteness, as well as being the most easily applicable based on publicly available information. The basic time periods of the study (2005–2007 for learning the models and 2008–2019 to assess the suitability of the predictions) are developed in a homogeneous regulatory environment, as far as the definition of failure is concerned.

3.2 Genetic Programming

3.2.1 Brief Commentary on Genetic Programming

Genetic programming (GP) (Koza, 1992) is an Evolutionary Computation (EC) technique that solves problems automatically without requiring the user to know or specify the form or structure of the solution in advance. At a more abstract level, it is a systematic, domain-independent method for getting computers to solve problems automatically based on high-level knowledge of what “needs to be done” (Poli et al., 2008).

EC methods simulate on computer the biological process of natural evolution, with natural selection as its driving force. A population of genotypes or individuals that encode solutions to a given problem is maintained. The solutions are evaluated and assigned a quality, *fitness*, based on how well they solve the problem. In general, these solutions are subjected to different “genetic operators”, which modify the genotypes of these solutions, operators such as crossover of genetic material between two individuals and mutation of part of the genetic material of a solution. The genetic operators are used to define new individuals and the selection operator is used to determine which individuals will become part of the next generation. The selection operator must respect Darwinian natural selection, in the sense that the higher the quality, the higher the probability of passing the genetic material of an individual to the next generation. This iterated process over different generations will obtain solutions that will progressively better solve the specific problem they encode. This general process is summarised in different algorithms in EC, such as the classical Genetic Algorithms, Evolutionary Strategies and Genetic Programming (Petrowski & Ben-Hamida, 2017).

What is distinctive in GP is that a population of computer “programs” evolves, programs that are often represented as decision trees (programs that correspond in our application to bankruptcy prediction models). In other words, generation by generation, GP transforms—stochastically—populations of programs into new and possibly better populations of programs. The steps and genetic operators of GP are detailed, for example, in Poli et al. (2008). GP’s ability to obtain programs automatically is what gives it great versatility and allows it to tackle problems not easily tackled by other evolutionary computation techniques (Petrowski & Ben-Hamida, 2017).

Not being the aim of this study the comparison of different Machine Learning (ML) techniques in the problem of business failure prediction, several features of GP are appropriate for the objective of obtaining prediction models with high and stable performance in the learning and after the learning period. First, GP makes no prior assumptions about the explanatory variables of a prediction model. Moreover, GP provides a straightforward interpretability of the optimised tree/program (Brabazon et al., 2020). GP also provides an automatic variable selection process, always tailored to each particular model. Finally, the ability to regulate in some way the complexity of the optimised tree or program, by parameterising its depth, the number of nodes or the functions that can be used in the search for a solution.

3.2.2 Genetic Programming Environment—Software

The software used to evolve prediction models by GP has been *HeuristicLab* (HL) (Wagner et al., 2014). This software can be downloaded from its website: <https://dev.heuristiclab.com/trac.fcgi/>. HeuristicLab was selected because it is an extensible and paradigm-independent optimisation environment that strongly abstracts the heuristic optimisation process, also with a detailed user interface (Wagner et al., 2014).

Briefly, the following process has been followed for the generation of each of the BPMs:

- The prediction problem is modelled in HeuristicLab as a Genetic Programming—Symbolic Classification problem.
- HeuristicLab is provided with a training set used in the progression of the evolutionary process (i.e., used to define the fitness of each solution). HL is also provided with a test set, consisting of different observations, in order for the software to apply the models obtained from the training set to this test set. The classification metrics for the selected solution/program will be provided with this test set.

3.3 Parameterisation of the Prediction Models

As already indicated, the objective is that the SMBPMs obtained present a high performance (similar to that of the MMBPMs) and interpretability for decision-making in their predictions over temporal horizons and economic environments other than those of the learning period. To this end, in line with other authors, some decisions have been taken, the most relevant of which are listed below.

3.3.1 Explanatory Variables

There is no consensus on the appropriate explanatory variables for a BPM. Although financial ratios are frequently used and have been shown to be effective in predicting business failure (du Jardin, 2009), many authors advocate the inclusion—usually together with financial ratios—of different explanatory variables. Some of the lines followed are shown below, without any intention of being exhaustive:

- Variables related to payment behaviour (Altman et al., 2015; Ciampi et al., 2020).
- Variables related to taxes (specifically the difference between accounting income and disposable income) (Noga & Schnader, 2013).
- Variables related to the company not related to financial ratios (e.g., age and audit reports) (Altman et al., 2015, 2020; Dakovic et al., 2010).
- Macroeconomic variables (Altman et al., 2015, 2020; Cybinski, 2001).

Several authors offer a more detailed view of the sets of explanatory variables used (Altman et al., 2015, 2020; du Jardin, 2017; Ratajczak et al., 2022). Generally, the selection of explanatory variables aims to improve the performance of the model. In our case, in addition to this objective, we intend to make the models interpretable, in the sense of facilitating decision-making. If among the explanatory variables there are variables on which the company has no capacity to act (e.g., the age of the company, the national unemployment rate, GDP growth, etc.), this interpretability is totally compromised, since the model does not facilitate decision-making as it is based on explanatory variables that do not allow the company to act on them (Altman et al., 2015, 2020; du Jardin, 2017; Ratajczak et al., 2022).

On the other hand, some authors (Beaver et al., 2005; Das et al., 2009; Tian et al., 2015) have found that some financial ratios constructed solely from accounting data contain relevant information about the risk of future failure. It has also been observed (Tian et al., 2015) that the importance of such ratios for predicting future failure, relative to market-based explanatory variables, increases as prediction horizons increase. Market variables are more useful in the short term than in the long term.

On the basis of the above (sufficiency of the financial variables) and some of the objectives of the study (interpretability for decision-making), the use of market variables is ruled out, so that the explanatory variables of the model are calculated from the data provided by the annual accounts of the companies and are mostly financial ratios.

The explanatory variables to be used were chosen on the basis of two criteria:

- The first criterion refers to the ratios selected on the basis of the relevance reported in the literature and their presence in the predictive models tested in previous work.
- The second criterion covers other variables obtained from the annual accounts and which refer to other aspects that are little used (e.g., variations in magnitudes or variations in ratios) or very infrequent or novel in this type of work (e.g., degree of decomposition of the balance sheet, those referring to productivity or those referring to fraud).

A number of works have been relevant in the above selection process, the main ones being—in no order of precedence—those of: du Jardin (2010), Bellovary et al. (2007), Altman and Sabato (2007), Altman et al. (2015), Yardeni et al. (2019), Beneish (1999) and Tian and Yu (2017).

To summarize, the initial set of explanatory variables can be grouped as follows:

- Liquidity and solvency: 18 ratios
- Financial structure: 14 ratios
- Profitability: 12 ratios
- Efficiency: 11 ratios
- Turnover: 7 ratios
- Variations in magnitudes: 3 ratios
- Contribution: 2 ratios

- Interest expenses: 5 ratios
- Size: 4 variables
- Growth: 1 ratio
- Changes in ratios: 2 ratios
- Degree of decomposition: 3 variables
- Productivity: 4 ratios
- Fraud: 11 variables

Thus, a total of 97 explanatory variables are used, which refer to a broad spectrum of aspects considered, a priori, to be relevant in business failure. It is necessary to highlight the large number of explanatory variables included in this study, which is much higher than that usually used in this type of studies. Moreover, in the case of GP, the ability to automatically (and intrinsically) select the relevant explanatory variables is an advantage over other ML techniques. That is, the selection pressure of the evolution of the final optimized programs/trees also involves an automatic determination of the relevant input variables to the prediction/classifier program.

3.3.2 Functions Used in the Evolved GP Programs

In this work, we have opted for the exclusive use of the arithmetic functions (addition, subtraction, multiplication and division) in the evolved GP programs (BPM models). The use of only the arithmetic functions has two additional effects on the results offered compared to the other options:

- Solutions are more interpretable, in the sense that the solution results provided by artificial intelligence can be better understood by human experts, in line with so-called explainable artificial intelligence (XAI).
- Solutions are simpler and this, as Finlay points out—cited by du Jardin and Séverin (2012)—, directly affects the stability of the predictive power of a solution over time, because the more complex a classifier is, the more often it must be re-estimated. This does not mean giving up the performance of the models. Balcaen and Ooghe (2006) conclude that simple models can gain significantly in classification accuracy compared to complicated models, due to the 80/20 Pareto rule and the law of diminishing returns.

3.3.3 Transformation of variables

The set of input variables will not be used with their original values, but will be transformed by standardising them according to the logistic distribution (Eq. 1). For this purpose, the mean and standard deviation of each of the variables in the period 2005–2007 will be used.

$$F_X(x; \mu, s) = \frac{1}{1 + e^{-(x-\mu)/s}} \quad (1)$$

In line with other authors (du Jardin, 2017; Nyitrai, 2019; Tascon et al., 2018), this transformation aims not only at standardisation, but also at quantifying how the financial health of the firm varies over time (by referring the values of the period 2008–2019 to the mean and standard deviation of the learning period 2005–2007).

3.3.4 Extension of the Training Period

We take the Laitinen (1991) idea of trying to capture the evolution of the explanatory variables by extending the range of the data used to create the model and by introducing time variations explicitly among these explanatory variables (e.g., among others, groups of variables such as variation in magnitudes, changes in ratios, degrees of decomposition, incorporate time variations of period t versus $t-1$). This idea is also pointed out by Shumway (2001) when the author comments about static models, indicating that the characteristics of most firms change from year to year. These (static models) ignore data on healthy firms that eventually go bankruptcy.

3.3.5 GP Parameterisation Overview

The most relevant parameters (using, where appropriate, the HL nomenclature) are listed in Table 1 together with their values/options. Some parameters are the usual values set in HeuristicLab (such as Solution Creator and Model Creator), while the others (such as Tournament window size, Population Size, Generations, Mutation

Table 1 GP parameters

GP parameter	Value/option
Explanatory variables	As outlined in Sect. 3.3.1
Variable transformation	Logistic distribution
Evaluator (fitness function)	Mean squared error (between predicted and real values)
Solution creator	Probabilistic tree creator
Symbolic expression tree grammar	Arithmetic functions
Maximum depth	10 (maximum depth of the tree/program)
Maximum length	100 (maximum number of nodes in the tree – symbolic classification model)
Maximum generations	100
Mutation probability	15%
Population size	1500
Selector	Tournament—Window size: 8 (used in mutation and crossover)
Positive class	1 (failed companies)
Elites	1 (only the best solution is retained between generations)
Crossover	Subtree Swapping Crossover at the crossing point)
Mutation	Multi Symbolic Expression Tree Manipulator (allows for different types of mutation)
Model creator	Accuracy Maximizing Thresholds (the returned solution is the one that uses as classification threshold the one that maximises the percentage of success in the training set)

Probability, Maximum Depth and Maximum Length in evolved trees) were experimentally selected to provide solutions with high classification performance.

This selection in these last parameters was performed with a standard/usual sweep of parameters in Evolutionary Computing. In this procedure, first a finite set of the values of each of the parameters is taken into account (e.g., discrete mutation rates between 1 and 50%, a low and a high rate for mutations). Second, for the sweep of the values in a parameter, the values of the other parameters are set to a standard or default value. Finally, the configuration of parameter values is chosen as the one that provides solutions with the highest classification performances according to the application interest (the highest normalized Gini coefficient over the estimated values using the test set, as detailed later in Sect. 4). Furthermore, given the stochasticity of the GP evolutionary process, a large number (1000) of independent GP runs (for each parameter configuration) are considered to determine the best solutions.

Figure 1, in the subsequent Sect. 5.3, includes an example of a program evolved according to all these aspects.

4 Methods of Obtaining MMBPMs and SMBPMs

To obtain each of the necessary temporal prediction models, an experiment is performed again with 1000 independent GP runs, given the stochasticity of the GP evolutionary process. From the 1000 evolved programs, a selection is made to choose the final evolved program corresponding to each of the prediction models considered in the study.

Specifically, in any GP experiment, the best models (evolved GP programs) correspond to those with the highest normalized Gini coefficient on the estimated values using the test set, after a filtering in which only solutions with *True Positive Rate* (TPR) + *True Negative Rate* (TNR) greater than a threshold are taken into account. This is intended to rule out—at least partially—those solutions with high AUC (Area Under the ROC Curve) thanks to its value in “extreme zones” in their ROC (Receiver Operating Characteristic) curve (i.e., in the extreme zones in the ROC Curve: high TPR and FPR—False Positive Rate, low TPR and FPR). On the contrary, the solutions that are not discarded are forced to go through a certain area of the ROC curve (at least one of its points), which can be called the area of interest. The final objective is to use the filter to detect—from the point of view of the work—the best solutions of each GP experiment. This approach is neither new nor strange in the medical field (Dodd & Pepe, 2003; McClish, 1989).

Three SMBPMs are considered in this study: SMBPM 1–3 (predicting failure at time interval 1–3 in the future), SMBPM 1–6 and SMBPM 1–9. Alternatively, 3 MMBPMs: MMBPM 1–3, MMBPM 1–6 and MMBPM 1–9 are also considered to predict failure at the same time intervals.

The MMBPMs are defined from the conditional probabilities of the annual BPMs that predict failure at $t+1$, $t+2$, ..., $t+9$. For example, MMBPM 1–3 uses 3 independent BPMs that predict failure at $t+1$, $t+2$ and $t+3$, which will allow to obtain the conditional probabilities (corresponding to the 3 BPMs) to define the MMBPM with prediction horizon 1–3. Similarly, MMBPM 1–6 uses 6 independent

BPMs that predict failure in the corresponding 6 years in the future and MMBPM 1–9 uses 9 independent BPMs that predict failure in their corresponding 9 years in the future. Consequently, 9 independent GP experiments are performed to obtain the 9 best BPMs to define the 3 MMBPMs considered (MMBPM 1–3, MMBPM 1–6, MMBPM 1–9).

In the case of SMBPMs, 3 independent SMBPMs are evolved, which predict failure at time intervals 1–3, 1–6 and 1–9. Therefore, 3 independent GP experiments are performed to obtain the best SMBPMs in these 3 time intervals considered (SMBPM 1–3, SMBPM 1–6, SMBPM 1–9).

From the annual models, the failure probabilities must be obtained so that—by means of the calculation of conditional probabilities—the MMBPMs with prediction horizons 1–3, 1–6 and 1–9 years prior to failure can be defined. To do this, for each of the annual prediction models (1, 2, ... and 9 years before failure), a probability distribution has to be fitted to the estimated values (output values of the model). To select the distribution that best fits the values estimated by each of the annual models, the results obtained by more than 30 distribution types are compared. The selection is made according to the maximum likelihood estimation method and the Kolmogorov–Smirnov test. In parallel, a Gaussian Mixture Model (GMM) is fitted to the estimated values of each of the annual models, the main characteristics of these models being the following:

- *Inference algorithm*: maximum expectation—MS (Dempster et al., 1977).
- *Selection criterion*: Bayesian information criterion—BIC (Schwarz, 1978).
- Mixture model with equal or different variance.
- *Number of classes*: 2.
- *A posteriori probability*: via the MAP rule (Maximum A Posteriori) (Bassett & Deride, 2019).

For each annual BPM, the fits obtained by the two methods (probability distribution adjustment and GMM posteriori probabilities) are compared and the best one is chosen. It should be noted that in the 9 annual prediction models, the probability distribution adjustment has shown better results than the probabilities obtained by means of GMM. In this way, the failure probabilities for each annual prediction horizon will finally be available, which will allow the calculation of the MMBPMs to be compared with the SMBPMs.

The time period in which the models are analysed in the learning period is 2005–2007, both inclusive. The number of observations used in the analysis in the learning period is as follows:

- Model 1–3:
 - *Training set*: 242 observations classified as failures and 242 observations classified as non-failures.
 - *Test set*: 242 observations classified as failures and 22,088 observations classified as non-failures.

- Model 1–6:
 - *Training set*: 677 observations classified as failures and 677 observations classified as non-failures.
 - *Test set*: 676 observations classified as failures and 21,653 observations classified as non-failures.
- Model 1–9:
 - *Training set*: 1013 observations classified as failures and 1013 observations classified as non-failures.
 - *Test set*: 1014 observations classified as failures and 21,317 observations classified as non-failures.

As indicated above, the training set drives the evolution of optimized programs (prediction models), while the test set is used to select the best final models.

5 Results

The results analyse the 6 considered multiperiod models:

- 3 MMBPMs (with prediction horizons 1–3, 1–6 and 1–9) obtained by means of conditional probabilities of the annual BPMs.
- 3 SMBPMs (with prediction horizons 1–3, 1–6 and 1–9 years prior to failure).

Two comparisons are performed: i) SMBPMs are compared with external references with data from the learning period. ii) The performance of SMBPMs is compared with that of MMBPMs. In the latter case, the comparison includes data after the learning period, using data corresponding to a wide period with different economic environments. Finally, SMBPMs have the important advantage of interpretability for decision-making and here two interpretability possibilities are considered for decision-making with evolved SMBPMs.

5.1 Performance of the Models in the Learning Period. Comparison with External References

It is not easy to relativise the performance of SMBPMs in the learning period largely because of the scarcity of external references (that use multiperiod models). Moreover, each external reference has used a different dataset and a different classifier, among the most relevant differences. Although the comparison is in all cases performed on the test set, this does not guarantee that the comparison is performed homogeneously since, for example, it may use different performance

measures (Chandrashekar & Sahin, 2014). In any case, it may be interesting to approximate whether or not the proposed models perform acceptably.

The AUC and the accuracy in the learning period (measured on the test set) of the proposed SMBPMs is as follows:

- *Model 1–3*: AUC 89.80%; Accuracy 98.92%.
- *Model 1–6*: AUC 84.06%; Accuracy 96.98%.
- *Model 1–9*: AUC 80.13%; Accuracy 95.47%.

As an external reference to compare and relativise the accuracy of SMBPMs, we will use the results of different models (MMBPMs) up to 5 years prior to failure:

- Luo et al. (2022) with parametric family learning through deep neural models.
 - Prediction horizon: 1–36 months prior to failure. Accuracy: 74.76%.
 - Prediction horizon: 1–60 months prior to failure. Accuracy: 65.32%.
- Orth (2013) with time-varying covariates and log-logistic model in conditional distribution.
 - Prediction horizon: 1–36 months prior to failure. Accuracy: 82.83%.
 - Prediction horizon: 1–60 months prior to failure. Accuracy: 79.31%.
- Duan et al. (2012) with a forward intensity approach and non-financial companies.
 - Prediction horizon: 1–36 months prior to failure. Accuracy: 66.98%.
- Duan and Fulop (2013) with the partially-conditioned forward intensity approach and non-financial companies.
 - Prediction horizon: 1–36 months prior to failure. Accuracy: 68.20%.

Note that the horizons in these external references are not completely coincident (with the intervals considered in the SMBPMs) and that they do not reach the 1–9 year horizon, for which no external multiperiod BPM references have been found. However, it is easy to see that the accuracy of SMBPMs is very high with respect to the benchmarks used.

Assessing and relativising the AUC of SMBPMs is more difficult since no external AUC references are available for MMBPMs. However, the AUC of SMBPMs can be relativised by comparing with the AUC of BPMs at different years prior to failure (which make predictions for a specific moment in time rather than a time interval). As an external reference we will use the synthesis compiled by Ratajczak et al. (2022). This work includes, among other sections, the analysis of the predictive capacity of different models up to 5 years prior to failure. The study by Ratajczak et al. (2022) contains references in which the capability of the models is measured in

terms of AUC. The study by Altman et al. (2015) is also used as a reference. Table 2 shows the AUC for BPMs in different years prior to failure.

Although this is not a comparison, but a relativisation, the performance of SMBPMs in terms of AUC (during the learning period and evaluated over the test set), presents a very high value. For example, the SMBPM with the longest prediction horizon (Model 1–9) presents an AUC of 80.13%, which is higher than that of any BPM that makes predictions 3 years before the failure.

5.2 Performance of SMBPMs and MMBPMs After the Learning Period

The performance of SMBPMs is now compared with that of MMBPMs. The comparison metric will be AUC and will be made on the basis of data after the learning period.

The time periods in which the models are analysed after the learning period are:

- *Model 1–3*: From 2008 to 2017, both inclusive. Therefore, for example: with the 2008 data, failure is predicted in the years 2009–2011, with the 2009 data, failure is predicted in the years 2010–2012, and with those of 2017 failure is predicted in the years 2018–2020.
- *Model 1–6*: From 2008 to 2014, both inclusive.
- *Model 1–9*: From 2008 to 2011, both inclusive.

These periods are the maximum possible based on the availability of data referring to failure in this study (until 2020) and present clearly different economic environments from that corresponding to the learning period (2005–2007). This, together with the breadth of these periods, means that the conclusions obtained from the results of the comparisons are much more reliable and well-founded than those obtained from a single comparison on the test set considering only the learning period.

The number of observations used in the analysis after the learning period is as follows:

- *Model 1–3*: In total there are 81,556 observations, of which 1533 are observations classified as failures (corresponding to observations from 1 to 3 years prior to failure, i.e., data on companies that fail within 1 to 3 fiscal years following the date of the observation) and 80,023 are observations classified as

Table 2 AUC for bankruptcy prediction models in different years prior to failure

	Years to failure					
	1 (%)	2 (%)	3 (%)	5 (%)	7 (%)	9 (%)
Altman et al. (2015)	88	82	78	76	73	69
Climent et al. (2019)	82	88	80			
du Jardin (2017)	83	80	79	77		
Tian et al. (2015)	84	73	68	59		

non-failures. This means that observations classified as failures account for 1.88% of all observations.

- *Model 1–6*: in total there are 56,756 observations, of which 2338 are observations classified as failed and 54,418 are observations classified as not failed.
- *Model 1–9*: in total there are 32,205 observations, of which 2093 are observations classified as failed and 30,112 are observations classified as not failed.

The figures are different because of two basic circumstances: i) the length of the temporal horizon of the post-learning period, which is different in each of the models and ii) basically referring to observations of failed firms, the absence of annual accounts as the end of a failure process approaches (Balcaen & Ooghe, 2006).

The following tables (Tables 3, 4 and 5) show the annual AUC of the SMBPMs and MMBPMs for each of the prediction horizons. Both multi-period alternatives are applied with the information data corresponding to fiscal years 2008 and later, analysing the AUC data in several consecutive fiscal years and after the period used in the training, also analysing the comparison between both alternatives.

In summary, the following can be observed:

- The average AUCs of SMBPMs are slightly lower than those of MMBPMs, although the differences are minimal (the largest difference occurs at the prediction horizon 1–3 years prior to failure and is, on average, -1.26%).
- This difference (always in favour of MMBPMs) decreases significantly as the prediction horizon increases (in the prediction horizon of 1–9 years prior to failure the difference is, on average, -0.20%). This is almost always the case in each of the years considered (the percentage difference in favour of MMBPMs in 2008 for a prediction horizon 1–3 is greater than the same for the 1–6 horizon, and the latter is greater than that for the 1–9 horizon). The only exception is 2009, where the prediction horizon 1–6 has the lowest difference.
- On average, there is no significant deterioration in the SMBPM performance (measured in terms of AUC) when SMBPMs are applied after the learning period. The AUCs in the learning period (measured on the test set) of SMBPMs are (as indicated above): Model 1–3: 89.80%; Model 1–6: 84.06% and Model 1–9: 80.13%. Consequently, in the prediction horizon 1–3 years prior to failure, there is a deterioration of 4.01% (considering the average AUC after the learning period shown in Table 3, 86.20%). Similarly, the deterioration is 0.25% on prediction horizon 1–6 and becomes an improvement of 2.45% on horizon 1–9.
- The AUCs obtained by the SMBPMs are stable over time. The highest Pearson's coefficient of variation is 2.26% on the prediction horizon 1–3 years prior to failure.

In many problems, including the prediction of business failure, regardless of the total AUC of a solution, it is relevant to analyse the AUC in a certain area of the ROC curve, since when evaluating a ROC curve, there are two areas that describe the behaviour of that solution in “extreme” situations:

Table 3 AUC per year—prediction horizon 1–3

	SMBPM1_3 (%)	MMBPM1_3 (%)	% variation
2008	88.5560	89.4848	–1.04
2009	86.3706	88.7280	–2.66
2010	85.2801	86.6378	–1.57
2011	83.8990	86.0003	–2.44
2012	84.2928	86.7106	–2.79
2013	84.1778	88.4229	–4.80
2014	85.4971	86.4364	–1.09
2015	85.6623	86.4512	–0.91
2016	88.5845	86.0889	2.90
2017	89.6638	88.0271	1.86
Average (years)	86.1984	87.2988	–1.26
Pearson's coefficient of variation	2.2589	1.3558	

Table 4 AUC per year—prediction horizon 1–6

	SMBPM1_6 (%)	MMBPM1_6 (%)	% variation
2008	83.6049	83.8420	–0.28
2009	84.9926	85.2670	–0.32
2010	85.0496	85.4767	–0.50
2011	82.9702	84.3972	–1.69
2012	83.2531	84.3513	–1.30
2013	83.5879	84.7677	–1.39
2014	83.4995	83.0984	0.48
Average (years)	83.8511	84.4572	–0.72
Pearson's coefficient of variation	0.9157	0.8985	

Table 5 AUC per year—prediction horizon 1–9

	SMBPM1_9 (%)	MMBPM1_9 (%)	% variation
2008	81.4053	81.5666	–0.20
2009	82.4991	83.1306	–0.76
2010	83.4991	82.9969	0.61
2011	80.9419	81.3191	–0.46
Average (years)	82.0863	82.2533	–0.20
Pearson's coefficient of variation	1.2091	0.9927	

- The area on the right, where the True Positive Rate (TPR) is high, accompanied by a high False Positive Rate (FPR). In this area, the acceptance rate of the model (when the model labels an observation as negative, rate measured as (True Negatives + False Negatives) / Total of samples)) is at low values, reaching 0% when both TPR and FPR are 100%.
- The area on the left, where low TPR and low FPR coexist. In this case, the acceptance rate is very high (reaching 100% if TPR and FPR are 0%), which makes the application of the model useless because it is not discriminatory.

In the case of the prediction of business failure, it will generally be interesting to focus attention on the central areas of AUC, excluding the aforementioned extreme zones. The following tables (Tables 6, 7 and 8) show the behaviour of the different SMBPMs and MMBPMs for each of the prediction horizons and years of analysis at different intervals in the central part of the ROC curve. The intervals analysed have been set completely arbitrarily and in terms of the model acceptance rate. AUC is measured in terms of TPR and FPR, but to each level of TPR and FPR corresponds a model acceptance rate, since the complementary of TPR is False Negative Rate and the complementary of FPR is True Negative Rate. The intervals analysed are as follows:

- *Interval 1:* Model acceptance rate between 90 and 10%.
- *Interval 2:* Model acceptance rate between 80 and 20%.

In these tables, the best average (years) and maximum values for each of the 2 intervals are highlighted in bold. It is clear from these tables that the slight difference in favour of MMBPMs is diluted as soon as different areas of the curve are analysed at different prediction horizons, without it being possible to draw a

Table 6 AUC per year and interval of acceptance rate—prediction horizon 1–3

	SMBPM1_3		MMBPM1_3	
	Interval 1 (%)	Interval 2 (%)	Interval 1 (%)	Interval 2 (%)
2008	54.1394	47.6832	56.9370	50.7718
2009	63.5194	52.7493	62.3928	55.7457
2010	67.8708	55.2896	71.2170	42.1746
2011	68.6251	48.2818	70.5411	45.0194
2012	68.4094	37.9707	54.5264	48.3229
2013	62.3879	55.5707	73.6249	51.8310
2014	63.0031	54.6704	63.0900	55.5595
2015	61.5981	54.6016	54.6610	47.9323
2016	55.3193	48.5251	55.2143	52.4807
2017	69.3008	41.9792	67.2778	38.0501
Average (years)	63.4173	49.7322	62.9482	48.7888

Table 7 AUC per year and interval of acceptance rate—prediction horizon 1–6

	SMBPM1_6		MMBPM1_6	
	Interval 1 (%)	Interval 2 (%)	Interval 1 (%)	Interval 2 (%)
2008	68.3303	54.7500	62.5770	52.2689
2009	68.9857	53.7986	70.0183	53.6640
2010	67.1858	47.9664	69.2234	55.2308
2011	64.4916	53.7080	70.1329	52.2445
2012	61.7091	52.4025	58.5920	52.9089
2013	59.9845	54.5609	71.2851	48.5525
2014	61.5765	47.5253	57.7694	51.6197
Average (years)	64.6091	52.1017	65.6569	52.3556

Table 8 AUC per year and interval of acceptance rate—prediction horizon 1–9

	SMBPM1_9		MMBPM1_9	
	Interval 1 (%)	Interval 2 (%)	Interval 1 (%)	Interval 2 (%)
2008	68.2607	52.5729	67.0689	51.9052
2009	69.7080	53.3848	61.3066	54.2810
2010	66.2388	51.8106	66.3054	53.4941
2011	67.3729	51.5698	67.4866	52.0316
Average (years)	67.8951	52.3345	65.5419	52.9280

conclusion about the superiority (in terms of performance) of any of the ways of defining multiperiod BPMs.

5.3 Interpretability of the Models

This section examines the interpretability of SMBPMs and their ability to facilitate informed decision-making.

There is no single definition of interpretability. A simple definition is given by Kim et al. (2016) when the authors indicate that interpretability is the degree to which a human being can consistently predict the outcome of the model. To increase the confidence and transparency of the models, Miller (2019) indicates two complementary approaches:

- Generate decisions in which one of the criteria taken into account during the calculation is how well a human could understand the decisions in the given context, which is often referred to as explainability or interpretability.
- Explicitly explaining decisions to people, which is called explanation. In our case, the explanation of the predictions of the predictors/classifiers (generally usable for individual predictions, i.e., referred to an observation).



Fig. 1 SMBPM 1–6 solution in hierarchical format

The first approach has been considered with the choice of the tool itself (GP). GP provides a straightforward interpretability of the optimised tree/program. As Brabazon et al. (2020) indicate, GP can provide human-readable solutions.

In our view, the interpretability of the proposed multiperiod BPMs—SMBPMs—should not be limited only to the ability to understand how they work, but should encompass explanation—basically of individual predictions—in two main ways: 1) explaining how a situation has been reached (projection into the past) and 2) facilitating the identification of what actions to take in the future, from a present situation and with a given objective, e.g., avoiding prediction of failure (projection into the future).

The full development of this approach is completely beyond the scope of this study, but the basic lines are outlined to show that the SMBPMs analysed are interpretable in the above sense. The approach proposed in this study is limited exclusively to building interpretation models with the following features:

- Post hoc (the model will be analysed after the training period).
- Specific to each model. The interpretation of the GP solution for each of the models is specific to each model, since the explanatory variables may be different and their relationship will also be different from one model to another.
- Particular to each prediction. The method of interpretation will explain an individual prediction, although generalizations can be made by aggregation of individual predictions.

One of the peculiarities of BPMs whose explanatory variables are based on financial statements is that these explanatory variables, although not necessarily correlated (in the statistical sense), can do have relationships between them. For example, a reduction in fixed liabilities may result in a reduction in assets (if the

source of financing is not replaced by another of the same amount). This prevents the marginal analysis of the explanatory variables, since it is not in accordance with reality. This is the reason for using an interpretation method based on the joint variations of explanatory variables. On the other hand, it should be remembered that the information available is limited to the large magnitudes of the financial statements and are positions at the end of each financial period. This will make interpretability coarse-grained for decision-making, but it will be sufficient to explain the model in adequate detail, although not in full detail.

The proposed method of interpretation and its scope are briefly outlined below. The model used as an example is the one corresponding to a prediction horizon 1–6 years prior to failure. The solution of SMBPM 1–6 in hierarchical format is shown Fig. 1.

In Fig. 1, the input variables are denoted in HL as r112_log_XX, where “XX” refers to the specific input variable (from the 97 indicated in Sect. 3.3.1) and “log” corresponds to the logistic distribution standardisation used in the variables. These explanatory variables used in the solution (SMBPM 1–6) are shown in Table 9.

Briefly, the method of interpretation will be based on analysing the model under the following conditions in order to facilitate interpretability and explanation:

- The elementary magnitudes that define the explanatory variables will be analysed (e.g.: Current Liabilities, Cash and Total Assets will be analysed, instead of the ratio (Current Liabilities – Cash) / Total Assets.
- Untransformed elementary magnitudes (in monetary units, employees, etc.) will be analysed. The model calculates explanatory variables, limits them and transforms them according to the logistic distribution.

Table 9 Explanatory variables of SMBPM 1–6

Name	Definition
r112_log_5	Current liabilities / Total assets
r112_log_24	Total debt / Total assets
r112_log_27	(Current liabilities – Cash) / Total assets
r112_log_40	Operating income / Total assets
r112_log_42	Funds from operations / Total liabilities
r112_log_43	Operating result / Total assets
r112_log_59	Accounts payable / Total sales
r112_log_62	Current liabilities / Total sales
r112_log_63	Financial expenses / Total sales
r112_log_65	Interest expense percent of equity
r112_log_67	Interest expense percent of net income
r112_log_69	Financial expenses / Total debt
r112_log_89	Log (Total assets)
r112_log_96	Number of employees

- Simple and homogeneous scenarios (joint variations of explanatory variables) will be considered. These scenarios are not the only possible ones nor are they necessarily disjoint with each other.

5.3.1 Explanation of the Evolution from the Past to the Present

The first challenge at the interpretability level is to explain how from a starting situation in the past the present situation (or another one in the past after the first one) has been reached. For illustrative purposes only and in a simplified way, a company is taken to exemplify the process: in this case, the one labelled #1694 and two instants in time (e.g., 2008 and 2010), to see what kind of explanation should be required from the model.

With data from company #1694 in 2008, SMBPM 1–6 prediction model offered an estimated value of 0.4018, which was in the 56th percentile of the total estimated values corresponding to the 7822 observations in 2008. In this case, and given the structure of the model, a lower estimated value indicates a better relative position (lower probability of failure) and a lower percentile.

The data for company #1694 in 2010 brought the value estimated by SMBPM 1–6 to 0.2323, which was in the 36th percentile according to the 2008 intervals. Percentiles are used as an indicator because it allows the estimated value to be relativised. On the other hand, it is chosen to maintain the limits of the percentiles in those corresponding to 2008, which allows to relativise the estimated value with respect to a common base of analysis—2008—with which the comparison is made. There is the option to calculate the 2010 percentile taking into account the estimated values of the model in 2010, but this does not provide information for comparison with the position in 2008 (e.g., it is possible to maintain an estimated value of an observation the same in 2008 and 2010, but the respective percentiles could vary if we calculate them with respect to 2008 and 2010 if the rest of the observations vary in estimated value). Therefore, percentile analysis is used to support interpretation.

We now proceed to analyse the variation from the initial position in 2008 to that corresponding to 2010. To do this, basic techniques of causal analysis are used that help to know what happens to a variable (estimated value) when others are changed (basic magnitudes).

The first step might be to disaggregate the total change in the estimated value according to the main groups of the available annual financial statements (variations related to balance sheet, variations related to employees, variations related to sales and other variations). The objective is to focus on the major causes of variation. Each of these groups is actually the aggregation of a set of more detailed variations. For example, variations related to balance sheet comprise changes in fixed assets, fixed assets depreciation rate, total debt, cost of the debt and others (e.g., balance sheet rates).

Now we take, for example, the case on variations related to employees (which includes the variation in the number of employees and the variation in the cost of employees). To calculate the impact on the estimated value of this change relative to employees, the value of the number of employees and the cost per employee of 2008 will be replaced by those corresponding to 2010, keeping the rest of the concepts in

the values of 2008. This substitution leads to a new estimated value. The variation of the estimated value obtained after replacement with respect to the estimated value of 2008 is made in two different ways: i) in percentage of variation with respect to the estimated value of 2008 and ii) obtaining the percentile of the estimated value with the substitution and calculating its difference in percentage points with respect to the corresponding value estimated in 2008—prior to the substitution—. The summary of the variation by large groups is shown in Table 10.

Table 10 includes as a reference the row “Total”, which corresponds to the variation of the estimated value (the output of the model of prediction) from the position in 2008 (0.4018) to that corresponding to 2010 (0.2323) in percentage (−42.19%) and percentile (from 56th in 2008 to 36th in 2010). Remember that, given the structure of the model, a lower estimated value indicates a better relative position (lower probability of failure). Also, in Table 10 the joint variations of the above groups have been ignored, this together with the fact that the proposed GP models are not linear (so they are not subject to the principle of superposition, which means that the behaviour of the system is not expressible as the sum of the behaviours of the descriptors), causes that the sum of the variation of the different groups does not coincide with the total variation. In any case, the main causes are quite clear. The main cause of the improvement in the estimated value has been changes in the balance sheet.

Continuing with the same example, the variations in balance sheet can be analysed up to the level that the available information allows. In more detail, the changes in balance sheet could be broken down as follows:

- Variation in fixed assets.
- Variation in the depreciation rate of fixed assets.
- Variation in total debt.
- Variation in the cost of debt.
- Other variations on the balance sheet (current assets and their breakdown, equity, etc.).
- Other balance rates.

The analysis can be made more precise by checking the data in the annual accounts. For instance, following the previous example of the company labelled #1694, it can be concluded that the improvement generated has its origin in a

Table 10 Explanation of evolution from past to present—analysis of the main variations of groups of variables

	% variation estimated value	Percentile variation
Variations related to balance sheet	−39.92	−19
Variations related to employees	−7.44	−4
Variations related to sales	4.97	2
Other variations	0.85	0
Total	−42.19	−20

decrease in the size of the company in the amount of 15,601.63 thousand euros (most of which—14,944.91 thousand euros—corresponds to indebtedness). Based on the available breakdown of the data, it can be seen that this decrease is reflected in a lower value of property, plant and equipment (−7581.15 thousand euros), a decrease in stocks (−2154.07) and debtors (−8023.05) and an increase in other liquid assets (2156.64). This, together with the variations in amortisation rates and costs of debt, justifies 19 points of negative variation in the percentile (with respect to the 20 points of variation of the reference “Total”, as shown in Table 10) or the 39.92% decrease in the estimated value (with respect to the 42.19% decrease considering the Total variation shown in Table 10).

5.3.2 Explanation of Future Actions

In this section we will now exemplify the decision-making process by taking the data of another company (e.g., the one labelled #17) and the 2014 data. The application of SMBPM 1–6 to such data provides an estimated value of 0.2284 which is in the 44th percentile of the estimated values corresponding to the 8289 observations of 2014. We will now consider changes in variables or sets of related variables grouped into “scenarios” to analyse the change in that percentile.

A number of simple scenarios (which incorporate the variation of a “central variable” and a set of elementary magnitudes that vary with it) can be defined and the marginal impact of each of them analysed. Note that this scenario definition is independent of the previous variable group definition. In this way, future opportunities to move away from failure as well as future threats to the company can be recognized. The scenarios considered in this study are not exhaustive, but they are common enough to provide a clear vision of how to generalize interpretability for future decision-making.

The percentages of variation of the main variable of each scenario have been chosen for illustrative purposes, to capture a wide range of variation. A key point to note is that each scenario involves the joint variation of several variables. Take, for example, the scenario called “Restructuring of total indebtedness (reduction of current liabilities and transfer to long term liabilities). The restructuring limit will be the amount of the balance of financial debts” and considered in the results shown in Table 11. The financial debts (which is a subset of current liabilities or short term liabilities) would be the central variable. In this example, this scenario implies the reduction of current debts (via reduction of financial debts) and the increase of long-term liabilities, while the rest of the variables remain fixed, including the value of the cost of the debt. If such a restructuring is estimated to modify the cost of debt, this cost of debt should also change. The idea is that a scenario captures variations in a set of variables related to each other by the scenario assumptions, while the rest of the variables remain constant.

For illustrative purposes only, the data included in Tables 11 and 12 show the variation in the percentile position when applying SMBPM 1–6 to company #17 with 2014 data, when the central variable is changed in a given percentage (note that the variation of a central variable has effects on the rest of the variables associated with that scenario). The tables show the results (variation in the percentile position)

Table 11 Explanation of future actions—negative percentage variations in different scenarios

Percentile variation in the predicted value	Percentage of variation of the central variable				
	- 50,00%	- 25,00%	- 10,00%	- 5,00%	- 1,00%
Reduction of total indebtedness by sale of fixed assets that are considered unnecessary for results or unproductive (applied first to reduce long term debt). The maximum reduction is the minimum value of the fixed assets considered non-productive and the sum of financial debt	0	0	0	0	0
Restructuring of total indebtedness (reduction of current liabilities and transfer to long-term liabilities). The restructuring limit will be the amount of the balance of financial debts	0	0	0	0	0
Reduction of long-term liabilities by increase of own funds. The limit of the reduction will be the amount of bank long term liabilities	-1	0	0	0	0
Reduction of the number of employees without affecting the operation. The reduction limit will be the number of employees considered non-productive	-12	-6	-3	-1	0

Table 12 Explanation of future actions—positive percentage variations in different scenarios

Percentile variation in the predicted value	Percentage of variation of the central variable				
	1,00%	5,00%	10,00%	25,00%	50,00%
Increase in property, plant and equipment (considered as a non-productive asset) with increase in long-term liabilities	0	0	0	0	0
Increase in property, plant and equipment (considered as a productive asset) with increase in long-term liabilities	-2	-10	-18	-34	-38
Increasing the number of employees without affecting production	1	5	9	23	40
Increasing the number of employees affecting production	-1	-7	-12	-23	-29

when negative variations (in the central variable) are applied in 4 different scenarios (Table 11) and positive variations are applied in 3 other scenarios (Table 12), all of them chosen arbitrarily. Moreover, these variations of the central variable of each scenario (e.g. total indebtedness, financial debts, bank long-term liabilities, etc.) are considered with different percentages of variation (values corresponding to each column in the tables).

It can be observed that the improvement (reduction) in the estimated value (decrease in the percentile) would be through one of the following alternatives:

- Reduce the number of employees without affecting the operation (operating income).
- Investing in property, plant and equipment (as productive assets) with an increase in long-term liabilities.
- Increase the number of employees resulting in an increase of operation (operating income).

On the contrary, the increase in the number of employees without leading to an improvement in operating income becomes the cause that can most worsen the estimated value (increase the estimated value and the percentile).

It should be remembered that the proposed interpretability approach is local and explains only individual predictions (for a particular company in a particular financial year). These explanations are not, to any degree, generalizable either to other companies with observations in the same year, nor to the same company with data in other years, etc.

6 Discussion and Conclusions

This study analyses the suitability of SMBPMs as an alternative to MMBPMs with three different prediction horizons (multiperiod models 1–3, 1–6 and 1–9). The study focuses on analysing the behaviour of the models after the learning period with the following perspectives:

- Performance.
- Interpretability for decision-making.

The main conclusions are, without priority, the following:

- Multiperiod BPMs for a given time horizon can be obtained directly and without resorting to obtaining intermediate models.

The SMBPMs presented (Model 1–3, Model 1–6 and Model 1–9) are obtained without the need for intermediate models and with no appreciable decrease of performance compared to models obtained on the basis of conditional probability (MMBPMs). Additionally, the SMBPMs obtained with GP are perfectly interpretable and allow future decision-making.

If, for example, Model 1–6 is considered, this model would allow us to calculate a probability of failure in the 1–6 years range. If this probability is calculated on the basis of the conditional probability in exercises 1, 2, 3, ..., 6, it involves fitting six probability functions, instead of one, with the risk of error that this may entail. On the other hand, the interpretability of a joint and/or conditional probability of 6 exercises would most likely suffer severely due to the difficulty in locating the relevant variables in the global prediction horizon 1–6.

- Multiperiod BPMs with long-term prediction horizons are possible.

From the analysis of the degree of deterioration of the AUC after the learning period, it can be concluded that the results question the idea of deterioration of the performance produced by the use of the model after the learning period. Table 13 summarizes the AUC results of the SMBPMs for the three prediction horizons considered (previously detailed). The AUCs obtained in the different prediction horizons do not show significant differences with respect to those achieved in the learning period.

Table 13 Average (years) of the AUC of the proposed SMBPMs

	In the learning period (measured on the test set) (%)	After the learning period (%)	% variation
SMBPM 1–3	89.80	86.20	– 4.01
SMBPM 1–6	84.06	83.85	– 0.25
SMBPM 1–9	80.13	82.09	2.45

This fact is more significant because the comparison is performed with respect to average AUCs obtained after the learning period, which are calculated on the basis of an average with wide observation periods. These periods range from 4 (Model 1–9) to 10 (Model 1–3) financial years. On the other hand, the AUCs after the learning period not only do not vary excessively with respect to those of the training period, but remain stable over time (as shown above in Tables 3–5).

- SMBPMs enable justified decision-making.

The interpretability of models is particularly useful when it can explain and inform future decisions aimed at counteracting the prediction of the BPM. This aspect allows a nuance to be introduced into the “black box” label of the proposed models. Consider, for example, a financial institution. The estimated value of company #17 in 2014 and its position (44th percentile, example considered in the previous Sect. 5.3.2.) would allow:

- Decide whether to take risks—or not—in the company (depending on the acceptance rate set by the financial institution).
- If the risk has already been assumed, the estimated value will serve to:
 - Detect warning signals (e.g., increase in the estimated value—probability of failure—since the date of risk assumption) and know the reasons for this change.
 - Design corrective measures, which can be proposed to the company (e.g.: propose to the company to reduce non-productive employees, offer financing to increase size and operation, etc.).
 - Decide on proposals by the company. If the company presents a liability restructuring plan, a decision can be made on the basis that the model does not estimate that this would substantially change the estimated value (the percentile does not change). That is, it is not the solution to improve the estimated value. On the other hand, if the firm submits a plan to reduce the number of non-productive employees, on the basis that the model estimates that such a plan will improve the estimated value (decrease the percentile), it may decide to accept such a plan.

It is interesting to note here that Directive (EU) 2019/1023 of the European Parliament and of the Council (2019) of 20 June 2019 on restructuring and insolvency (EU-DRI), article 3, requires Member States to provide tools for early warning of insolvency. Consequently, these tools should serve as a warning to debtors at risk of insolvency so that they can take prompt action to avoid insolvency. This approach of the interpretability of SMBPMs allows not only for early warning, but also for the optimisation of corrective actions.

- The financial variables of the company are sufficient to obtain highly effective and interpretable multiperiod BPMs.

The proposed models are capable of making highly effective predictions in a stable and sustained way over long periods of time and with highly changing environmental conditions between them (as those observed in the period 2008–2019, the post-learning period analysed in this study). All this without resorting to explanatory variables other than those that can be obtained from the financial statements of the company. This, in addition to corroborating the idea put forward by Tian et al. (2015), allows the solutions to be interpretable with the proposals presented, by excluding variables over which the company has no capacity to act. The works of Beaver et al. (2005) and Das et al. (2009) point along the same lines of the adequacy of financial variables.

- GP is shown to be a suitable ML technique for obtaining multiperiod BPMs.

The choice of the technique, for implementing and learning the models, is not a trivial issue, although it is not the aim of this work. The suitability demonstrated by GP is given by the degree of compliance (by the models obtained with GP) with the proposed objectives of deterioration, stability, efficiency and interpretability. It is possible that other ML techniques can achieve similar results, but GP demonstrates its good performance for the proposed challenge of high predictive power, as evidenced by the accuracies obtained in the learning period by the SMBPMs at the different prediction horizons (using GP), which are much higher than those of the available external references (obtained with other techniques). Also, the AUC of SMBPMs is very high when relativised to the AUC of BPMs that make predictions at a specific point in time (not in a time interval). As shown in Sect. 5.1, for example, SMBPM 1–3 provides an AUC of 89.80%, while the maximum value in the external references considered in the comparison for BPMs 1 year prior to failure is 88%. In the case of SMBPM 1–6 and SMBPM 1–9, they provide an AUC of 84.06% and 80.13% versus the best value of 80% in BPMs 3 years prior to failure.

In summary, the present study shows a way of obtaining multiperiod BPMs (SMBPMs) which, without detriment to the performance compared to models obtained by joint probabilities (MMBPMs), presents a notable advantage: interpretability for decision-making. Moreover, the study addresses long-term prediction (1–9 years), with remarkable results (the AUC of SMBPMs for 1–9 years prior to failure in the training period is 80.13%, even higher than 80% of conventional models 3 years prior to failure). Although it has been shown how SMBPMs can be used in decision-making, a line of future development is to further develop this study of interpretability for decision-making, seeking a more detailed causal analysis that can be considered as standardised.

Author Contributions All authors contributed to the study conception and design. Material preparation and data collection were mainly performed by Ángel Beade. Analysis of the results and drafting of the manuscript were performed by all authors. All authors read and approved the final manuscript.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This study was funded by the Xunta de Galicia and the European Union (European Regional Development

Fund—Galicia 2014–2020 Program), with grants CITIC (ED431G 2019/01) and GPC ED431B 2022/33, and by the Spanish Ministry of Science and Innovation (project PID2020-116201 GB-I00).

Declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589–609. <https://doi.org/10.1111/j.1540-6261.1968.tb00843.x>
- Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E., & Suvas, A. (2015). Financial and non-financial variables as long-horizon predictors of bankruptcy. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2669668>
- Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K., & Suvas, A. (2020). A race for long horizon bankruptcy prediction. *Applied Economics*, 52(37), 4092–4111. <https://doi.org/10.1080/00036846.2020.1730762>
- Altman, E. I., & Sabato, G. (2007). Modelling credit risk for SMEs: Evidence from the U.S. market. *Abacus*, 43(3), 332–357. <https://doi.org/10.1111/j.1467-6281.2007.00234.x>
- Aziz, M. A., & Dar, H. A. (2006). Predicting corporate bankruptcy: Where we stand? *Corporate Governance: The International Journal of Business in Society*, 6, 18–33.
- Balcaen, S., & Ooghe, H. (2006). 35 years of studies on business failure: An overview of the classic statistical methodologies and their related problems. *The British Accounting Review*, 38(1), 63–93. <https://doi.org/10.1016/j.bar.2005.09.001>
- Basel Committee on Banking Supervision. (2004). Basel II: International convergence of capital measurement and capital standards: A revised framework. *Bank for International Settlements*. <https://www.bis.org/publ/bcbs128.pdf>
- Basel Committee on Banking Supervision. (2009). Guiding principles for the replacement of IAS 39. *Bank for International Settlements*. <https://www.bis.org/publ/bcbs161.pdf>
- Basel Committee on Banking Supervision. (2011). Basel III: A global regulatory framework for more resilient banks and banking systems. *Bank for International Settlements*. <https://www.bis.org/publ/bcbs189.pdf>
- Basel Committee on Banking Supervision. (2017). Basel III: Finalising post-crisis reforms. *Bank for International Settlements*. <https://www.bis.org/bcbs/publ/d424.htm>
- Bassett, R., & Deride, J. (2019). Maximum a posteriori estimators as a limit of Bayes estimators. *Mathematical Programming*, 174(1), 129–144. <https://doi.org/10.1007/s10107-018-1241-0>
- Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, 71–111. <https://doi.org/10.2307/2490171>
- Beaver, W. H., McNichols, M. F., & Rhie, J.-W. (2005). Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy. *Review of Accounting Studies*, 10(1), 93–122. <https://doi.org/10.1007/s11142-004-6341-9>
- Bellovary, J. L., Giacomino, D. E., & Akers, M. D. (2007). A review of bankruptcy prediction studies: 1930 to present. *Journal of Financial Education*, 33, 1–42.

- Beneish, M. D. (1999). The detection of earnings manipulation. *Financial Analysts Journal*, 55(5), 24. <https://doi.org/10.2469/faj.v55.n5.2296>
- Blümke, O. (2022). Multiperiod default probability forecasting. *Journal of Forecasting*, 41(4), 677–696. <https://doi.org/10.1002/for.2825>
- Brabazon, A., Kampouridis, M., & O'Neill, M. (2020). Applications of genetic programming to finance and economics: Past, present, future. *Genetic Programming and Evolvable Machines*, 21(1–2), 33–53. <https://doi.org/10.1007/s10710-019-09359-z>
- Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1), 16–28. <https://doi.org/10.1016/j.compeleceng.2013.11.024>
- Chava, S., & Jarrow, R. A. (2004). Bankruptcy prediction with industry effects. *Review of Finance*, 8(4), 537–569. <https://doi.org/10.1093/rof/8.4.537>
- Christensen, J., Hansen, E., & Lando, D. (2004). Confidence sets for continuous-time rating transition probabilities. *Journal of Banking & Finance*, 28(11), 2575–2602.
- Ciampi, F., Cillo, V., & Fiano, F. (2020). Combining Kohonen maps and prior payment behavior for small enterprise default prediction. *Small Business Economics*. <https://doi.org/10.1007/s11187-018-0117-2>
- Climent, F., Momparler, A., & Carmona, P. (2019). Anticipating bank distress in the Eurozone: An extreme gradient boosting approach. *Journal of Business Research*, 101, 885–896. <https://doi.org/10.1016/j.jbusres.2018.11.015>
- Cybinski, P. (2001). Description, explanation, prediction—the evolution of bankruptcy studies? *Managerial Finance*, 27(4), 29–44. <https://doi.org/10.1108/03074350110767123>
- Dakovic, R., Czado, C., & Berg, D. (2010). Bankruptcy prediction in Norway: A comparison study. *Applied Economics Letters*, 17(17), 1739–1746. <https://doi.org/10.1080/13504850903299594>
- Das, S. R., Hanouna, P., & Sarin, A. (2009). Accounting-based versus market-based cross-sectional models of CDS spreads. *Journal of Banking & Finance*, 33(4), 719–730. <https://doi.org/10.1016/j.jbankfin.2008.11.003>
- Deakin, E. B. (1972). A discriminant analysis of predictors of business failure. *Journal of Accounting Research*, 10(1), 167–179. <https://doi.org/10.2307/2490225>
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (methodological)*, 39(1), 1–22. <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>
- Dodd, L. E., & Pepe, M. S. (2003). Partial AUC estimation and regression. *Biometrics*, 59(3), 614–623.
- dos Reis, G., & Smith, G. (2018). Robust and consistent estimation of generators in credit risk. *Quantitative Finance*, 18(6), 983–1001. <https://doi.org/10.1080/14697688.2017.1383627>
- du Jardin, P. (2009). *Bankruptcy prediction models: How to choose the most relevant variables?* [MPRA Paper]. <https://mpra.ub.uni-muenchen.de/44380/>
- du Jardin, P. (2010). Predicting bankruptcy using neural networks and other classification methods: The influence of variable selection techniques on model accuracy. *Neurocomputing*, 73(10), 2047–2060. <https://doi.org/10.1016/j.neucom.2009.11.034>
- du Jardin, P. (2017). *Dynamics of firm financial evolution and bankruptcy prediction*. <https://www.edhec.edu/en/publications/dynamics-firm-financial-evolution-and-bankruptcy-prediction>
- du Jardin, P., & Séverin, E. (2012). Forecasting financial failure using a Kohonen map: A comparative study to improve model stability over time. *European Journal of Operational Research*, 221(2), 378–396. <https://doi.org/10.1016/j.ejor.2012.04.006>
- Duan, J.-C., & Fulop, A. (2013). Multiperiod corporate default prediction with the partially-conditioned forward intensity. Available at SSRN 2151174.
- Duan, J.-C., Sun, J., & Wang, T. (2012). Multiperiod corporate default prediction—A forward intensity approach. *Journal of Econometrics*, 170(1), 191–209. <https://doi.org/10.1016/j.jeconom.2012.05.002>
- Duffie, D., Saita, L., & Wang, K. (2007). Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 83(3), 635–665. <https://doi.org/10.1016/j.jfineco.2005.10.011>
- Edmister, R. O. (1972). An empirical test of financial ratio analysis for small business failure prediction. *The Journal of Financial and Quantitative Analysis*, 7(2), 1477. <https://doi.org/10.2307/2329929>
- European Parliament. (2019). *Directive (EU) 2019/1023 of the European Parliament and of the Council—Of 20 June 2019—On preventive restructuring frameworks, on discharge of debt and disqualifications, and on measures to increase the efficiency of procedures concerning restructuring, insolvency and discharge of debt, and amending Directive (EU) 2017/1132 (Directive on restructuring and insolvency)*, 38.

- Financial Accounting Standard Board (FASB). (2016). Financial Instruments—Credit Losses (topic 326). *FASB Accounting Standards Update No. 2016–13*.
- International Accounting Standards Board. (2014). IFRS 9 Financial instruments. *International Financial Reporting Standards*. <https://www.ifrs.org/issued-standards/list-of-standards/ifrs-9-financial-instruments/>
- Jarrow, R. A., Lando, D., & Turnbull, S. M. (1997). A Markov model for the term structure of credit risk spreads. *The Review of Financial Studies*, 10(2), 481–523. <https://doi.org/10.1093/rfs/10.2.481>
- Kim, B., Khanna, R., & Koyejo, O. (2016). Examples are not enough, learn to criticize! Criticism for interpretability. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2288–2296.
- Kim, S. Y., & Upneja, A. (2021). Majority voting ensemble with a decision trees for business failure prediction during economic downturns. *Journal of Innovation & Knowledge*, 6(2), 112–123. <https://doi.org/10.1016/j.jik.2021.01.001>
- Koza, J. R. (1992). *Genetic programming: On the programming of computers by means of natural selection*. The MIT Press.
- Laitinen, E. K. (1991). Financial ratios and different failure processes. *Journal of Business Finance & Accounting*, 18(5), 649–673. <https://doi.org/10.1111/j.1468-5957.1991.tb00231.x>
- Luo, W.-L., Lu, Y.-M., Yang, J.-H., Duan, J.-C., & Wang, C.-J. (2022). Multi-period corporate default prediction through neural parametric family learning. In *Proceedings of the 2022 SIAM International Conference on Data Mining (SDM)* (pp. 316–324). <https://doi.org/10.1137/1.9781611977172.36>
- Matenda, F. R., Sibanda, M., Chikodza, E., & Gumbo, V. (2021). Bankruptcy prediction for private firms in developing economies: A scoping review and guidance for future research. *Management Review Quarterly*. <https://doi.org/10.1007/s11301-021-00216-x>
- McClish, D. K. (1989). Analyzing a portion of the ROC curve. *Medical Decision Making: An International Journal of the Society for Medical Decision Making*, 9(3), 190–195. <https://doi.org/10.1177/0272989X8900900307>
- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1–38. <https://doi.org/10.1016/j.artint.2018.07.007>
- Muslim, M. A., & Dasril, Y. (2021). Company bankruptcy prediction framework based on the most influential features using XGBoost and stacking ensemble learning. *International Journal of Electrical and Computer Engineering (IJECE)*, 11(6), 5549. <https://doi.org/10.11591/ijece.v11i6.pp5549-5557>
- Noga, T. J., & Schnader, A. L. (2013). Book-tax differences as an indicator of financial distress. *Accounting Horizons*, 27(3), 469–489.
- Nyitrai, T. (2019). Dynamization of bankruptcy models via indicator variables. *Benchmarking-an International Journal*, 26(1), 317–332. <https://doi.org/10.1108/BIJ-03-2017-0052>
- Orth, W. (2013). Multi-period credit default prediction with time-varying covariates. *Journal of Empirical Finance*, 21, 214–222. <https://doi.org/10.1016/j.jempfin.2013.01.006>
- Petrowski, A., & Ben-Hamida, S. (2017). *Evolutionary algorithms*. John Wiley & Sons.
- Poli, R., Langdon, W. B., McPhee, N. F., & Koza, J. R. (2008). *A field guide to genetic programming*. Lulu Press.
- Ratajczak, P., Szutowski, D., & Szulcowska-Remi, A. (2022). *Long-term bankruptcy prediction. Systematic literature review* (SSRN Scholarly Paper No. 4054665). <https://doi.org/10.2139/ssrn.4054665>
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464. <https://doi.org/10.1214/aos/1176344136>
- Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business*, 74(1), 101–124. <https://doi.org/10.1086/209665>
- Tascon, M. T., Castaño, F. J., & Castro, P. (2018). A new tool for failure analysis in small firms: Frontiers of financial ratios based on percentile differences (PDFR). *Spanish Journal of Finance and Accounting-Revista Espanola De Financiacion Y Contabilida*, 47(4), 433–463. <https://doi.org/10.1080/02102412.2018.1468058>
- Tian, S., & Yu, Y. (2017). Financial ratios and bankruptcy predictions: An international evidence. *International Review of Economics & Finance*, 51, 510–526. <https://doi.org/10.1016/j.iref.2017.07.025>
- Tian, S., Yu, Y., & Guo, H. (2015). Variable selection and corporate bankruptcy forecasts. *Journal of Banking & Finance*, 52, 89–100.
- Traczynski, J. (2017). Firm default prediction: A Bayesian model-averaging approach. *Journal of Financial and Quantitative Analysis*, 52(3), 1211–1245. <https://doi.org/10.1017/S002210901700031X>

- Tsai, C.-F., Sue, K.-L., Hu, Y.-H., & Chiu, A. (2021). Combining feature selection, instance selection, and ensemble classification techniques for improved financial distress prediction. *Journal of Business Research*, *130*, 200–209. <https://doi.org/10.1016/j.jbusres.2021.03.018>
- Wagner, S., Kronberger, G., Beham, A., Kommenda, M., Scheibenpflug, A., Pitzer, E., Vonolfen, S., Kofler, M., Winkler, S., Dorfer, V., & Affenzeller, M. (2014). Architecture and design of the HeuristicLab optimization environment. In R. Klemous, J. Nikodem, W. Jacak, & Z. Chaczko (Eds.), *Advanced Methods and applications in computational intelligence* (pp. 197–261). Springer.
- Wilcox, J. W. (1973). A prediction of business failure using accounting data. *Journal of Accounting Research*, *11*, 163–179. <https://doi.org/10.2307/2490035>
- Yardeni, E., Abbot, J., & Quintana, M. (2019). *S&P 500 Financial ratios*. Yardeni Research Inc.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.