Expert Systems With Applications 53 (2016) 87–105

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A novel corporate credit rating system based on Student's-*t* hidden Markov models

Anastasios Petropoulos[a], Sotirios P. Chatzis[b,*], Stylianos Xanthopoulos[a]

[a] Department of Mathematics, University of Aegean, Karlovasi 83200, Greece
[b] Department of Electrical Engineering, Computer Engineering and Informatics, Cyprus University of Technology, Limassol 3036, Cyprus

ARTICLE INFO

Keywords:
Corporate credit rating
Hidden Markov model
Student's-*t* distribution
Expectation-maximization
Basel framework
Statistical machine learning

ABSTRACT

Corporate credit rating systems have been an integral part of expert decision making of financial institutions for the last four decades. They are embedded into the pricing function determining the interest rate of a loan contact, and play crucial role in the credit approval process. However, the currently employed intelligent systems are based on assumptions that completely ignore two key characteristics of financial data, namely their heavy-tailed actual distributions, and their time-series nature. These unrealistic assumptions definitely undermine the performance of the resulting corporate credit rating systems used to inform expert decisions. To address these shortcomings, in this work we propose a novel corporate credit rating system based on Student's-*t* hidden Markov models (SHMMs), which are a well-established method for modeling heavy-tailed time-series data: Under our approach, we use a properly selected set of financial ratios to perform credit scoring, which we model via SHMMs. We evaluate our method using a dataset pertaining to Greek corporations and SMEs; this dataset includes five-year financial data, and delinquency behavioral information. We perform extensive comparisons of the credit risk assessments obtained from our method with other models commonly used by financial institutions. As we show, our proposed system yields significantly more reliable predictions, offering a valuable new intelligent system to bank experts, to assist their decision making.

1. Introduction

In this work, we focus on the problem of*creating intelligent tools to assist credit scoring/rating of individual corporations*. In general, a credit scoring/rating system makes use of a statistical technique that combines and analyzes a series of account statement data to predict the future behavior of a company in terms of its ability to service its debt. The used account data are usually in the form of financial ratios, while the system-generated predictions are typically quantified as the likelihood of occurrence of a default event at some specific future time point. Corporate credit rating systems have been an integral part of financial institutions decision making for the last four decades. Nowadays their use is widespread, and gives the loans officers the ability to discriminate customers based on their risk profile and make informed decisions during all the stages of a loan lifecycle. Rating expert systems are embedded into the pricing function determining the interest rate of a loan contact, and play crucial role in the credit approval process.

The introduction of the Basel II framework (Basel Committee on Banking Supervision, 2005a), and its continuation in Basel III (Basel Committee on Banking Supervision, 2010), has triggered immense interest in intelligent credit rating systems research. Specifically, under these frameworks, banks have to implement appropriate rating systems for estimating the probability of a customer becoming delinquent; this is used in order to assess the borrower's creditworthiness, and estimate the capital requirements attached to the specific loan contact. Furthermore, in order to comply with the regulations set out in Basel II/Basel III, financial institutions have to satisfy the use test, i.e. use the corporate rating systems in all the phases of the customer relationship (credit granting, provisioning, pricing, collateral management, collections and write offs). Finally, under Basel II/Basel III, banks have to build their rating systems using state-of-the-art *statistical* methods, as this is expected to increase the accuracy of their risk discrimination procedures, thus guaranteeing their good financial performance in the years to come. As such, since the introduction of Basel II, banks have been consistently motivated to develop more accurate and robust intelligent systems to drive expert decisions, exploring new statistical techniques especially from the field of statistical machine learning. In the last decades, a plethora of alternative approaches have been

* Corresponding author. Tel.: +35725002041.
E-mail addresses: tasospetropoulos@gmail.com (A. Petropoulos), soteri0s@me.com (S.P. Chatzis), sxantho@aegean.gr (S. Xanthopoulos).

http://dx.doi.org/10.1016/j.eswa.2016.01.015

developed to address the problem of modeling the credit quality of a company, using both quantitative information (e.g., account statements) and qualitative information (e.g., other underwriting criteria, such as obligors market and sector indicators).

A first category of approaches belongs to the family of classical regression techniques. Altman (1968) used multiple linear discriminant analysis (LDA) to build a rating system for predicting corporate bankruptcies. They estimated a linear discriminant function using liquidity, profitability, leverage, solvency, and turnover financial ratios to estimate credit quality; they dubbed their approach as the Z-score model. One of the main drawbacks of this approach is its assumption that the modeled variables are normally distributed, which is hardly ever the case in real-world scenarios. As such, this method cannot effectively capture nonlinear relationships among the modeled variables, which is crucial for the performance of the credit rating system. In a similar vein, several studies have explored the utility of probit models (e.g, (Mizen & Tsoukas, 2012)) and linear regression models (e.g., Avery, Calem, & Canner (2004)). However, these models continue to suffer from the same drawbacks that plague LDA, namely their clear inability to capture non-linear dynamics, which are prevalent in financial ratio data (Petr & Gurný, 2013).

Logistic regression is another approach broadly used for building corporate rating systems. It was first used by Ohlson (1980) to predict corporate bankruptcy based on publicly available financial data pertaining to several enterprises (e.g., financial ratios). Logistic regression models employed in this context are essentially used to classify corporations into two distinct classes characterizing their credit risk (i.e., good or bad). Typically, a sigmoid likelihood function is used for modeling purposes to allow for capturing non-linearities and relaxing the normality assumption during model estimation (Kamstra, Kennedy, & Suan, 2001).

Decision trees comprise a further category of non-parametric methods used for developing credit rating systems. Decision trees are models that consist of a set of nodes, corresponding to the modeled explanatory variables, and split conditions based on a hierarchical selection of the modeled explanatory variables. Two well-known algorithms in this field are the Chi-squared Automatic Interaction Detector (CHAID) (GV, 1978) and CART (Breiman, Friedman, Stone, & Olshen, 1984) techniques. Decision trees offer simplicity and flexibility in the employed modeling assumptions, while also allowing for easy visualization of the learned modeling strategies (obtained after training). On the negative side, the entailed variable discretization performed by these models results in potential loss of significant information, as well as overfitting proneness.

Another popular class of statistical models used for credit rating is hazard rate models. These models extend the time horizon of a rating system, by looking at the probability of default during the life cycle of the examined loan or portfolio (Chava & Jarrow, 2004; Shumway, 2001). To achieve this, hazard models explicitly model a survival function for the behavior of an examined borrower. Cox Proportional hazard models are one popular instance of this type of models (DR, 1972); it is based on the assumption that the covariates affecting the default rate are multiplicatively related to the hazard rate function (Im, Apley, Qi, & Shan, 2012).

Machine learning techniques have also been successfully applied to enhance the capabilities of conventional corporate credit scoring systems in several studies. Among such works, feedforward neural networks (FNNs) constitute the most commonly used machine learning method in the context of corporate credit rating systems (e.g., Zhao et al., 2015). Their successful application in the context of corporate credit rating is basically due to their nonlinear and non-Gaussian modeling assumptions, and their capability to capture dependencies between assets. On the negative side, the notorious proneness of FNNs to overfitting (and, thus, their limited generalization capacity), their need of tedious cross-validation to perform hyperparameter selection (e.g., network size selection), along with their black-box nature that hinders intuitive visualization of the obtained results, limit their potential appeal to the financial community. Other researchers have considered using support vector machines (SVMs) (Vapnik, 1998) to effect the credit rating task. Indeed, a significant number of studies published in the last decade have shown that SVMs outperform FNNs in credit rating scenarios (Chen & Li, 2014; Chen & Shih, 2006; Danenas & Garsva, 2015; Harris, 2015; Huang, 2009; Wang & Ma, 2012; Yi & Zhu, 2015), while reducing the possibility of overfitting, and alleviating the need of tedious cross-validation for the purpose of appropriate hyperparameter selection. On the negative side, SVMs also constitute black-box models, thus limiting their potential of offering deeper intuitions and visualizations regarding the obtained results of their modeling and inference procedure. A Bayesian inference-based analogous to SVMs, namely Gaussian processes, have also been considered by Huang (2011). A drawback of this approach is its high computational complexity, which is cubic to the number of available data points, combined with the assumption of normally distributed data, which is clearly unrealistic, as we have already explained. Finally, Random Forests (RFs) is another type of methods that has recently garnered attention by researchers working in the field of corporate credit rating. This sophisticated technique was introduced in (Breiman, 2001), while one successful application of RFs to the problem of corporate credit rating can be found in (Yeh, Lin, & Hsu, 2012).

Despite these considerable advances, all *existing work on intelligent corporate credit rating systems fails to address and to appropriately take into account two key characteristics of financial data: (i) their heavy-tailed nature; and (ii) the entailed temporal dynamics/evolution that characterize the financial behavior of corporate entities.*To resolve these issues, in this paper we develop a novel holistic corporate credit scoring system, that addresses all the parts of the modeling pipeline, from financial ratio time-series selection and preprocessing, to selection of appropriate time-series modeling techniques, and information fusion strategies used to obtain the final credit scores. At the heart of the proposed system lies a novel financial data modeling scheme, capable of modeling data with heavy-tailed nature and intricate temporal dynamics (typical in the field of finance), based on *Student's-t hidden Markov models (SHMMs)* (Chatzis, Kosmopoulos, & Varvarigou, 2009).

Our developed expert system constitutes an intricate data processing pipeline comprising a financial data preprocessing stage, developed on the basis of expert knowledge, and a core modeling stage, where SHMMs are used to capture salient temporal patterns in the modeled time-series that are associated with different credit risk scores. To perform modeling and prediction, our approach utilizes appropriate financial ratio time-series, based on the assumption that financial ratios carry all the information necessary to describe and predict the internal state of a company. Specifically, we use five-year historical data of financial ratios, that provide adequate insights on how profitable an examined company is, what the trends are, and how much risk is embedded in its business models. We fit distinct SHMMs to each one of the modeled financial ratios, and obtain separate credit scores from each one of them. Eventually, we train one final information fusion layer that combines the outputs of the individual SHMMs under a weighted linear combination scheme, to generate the final predictions obtained by our system. Parameter optimization of this final information fusion layer is performed by means of a simple yet effective genetic algorithm (GA) (Deb, Agrawal, Pratap, & Meyarivan, 2000).

SHMMs are a successful machine learning technique for modeling data with *temporal dynamics* (i.e., time-series data), that may contain a number of *outliers* and related artifacts in the available training datasets. As such, SHMMs arise as a natural selection for effecting the task of modeling financial ratio data, which entail strong temporal dependencies, while also being quite likely to comprise significant proportions of outliers. This *key* modeling selection of our approach is in *stark*contrast to the machine learning methods used in the context of existing corporate credit scoring systems: existing approaches are based on machine learning models that *neither are capable of capturing temporal dependencies in the modeled data, nor can effectively handle outliers* in their training datasets. *Due to these significant modeling advantages, our approach is expected to yield much better discriminative performance compared to the intelligent tools currently used by banking experts.*

The remainder of the paper is organized as follows. In Section 2, we provide a brief overview of related work dealing with applications of hidden Markov models (HMMs) to credit risk prediction, and explain the differences between our novel approach and the existing corpus of works. In Section 3, we provide a concise introduction to HMMs, focusing on the case of SHMMs and their training and inference algorithms under the maximum-likelihood framework. In Section 4, we introduce our proposed system, and elaborate on the used data selection and preprocessing schemes, the adopted modeling assumptions and strategies, and the associated training and inference algorithms. In Section 5, we perform the experimental evaluation of our approach: Initially, we elaborate on our experimental setup, and provide details regarding our implementation of the considered alternative methods that we evaluate in parallel to our approach (for comparative purposes). Further, we present our empirical results, analyzing the performance of our proposed SHMM-based corporate credit scoring model, and comparing its performance to the considered state-of-the-art competitors. Finally, in the concluding section, we highlight the performance advantages of our approach, we outline possible limitations of our framework, and discuss areas for future enhancements and research.

2. Existing applications of HMMs to credit risk assessment

HMMs constitute a rather popular method in financial literature. However, up to now their applications to risk assessment have been mostly limited to quantifying risk on *portfolio* level, as opposed to *individualcompany* level, which is the aim of this work. Specifically, HMMs first appeared in the financial literature in (Thomas, Allen, & Morkel-Kingsbury, 2002); therein, the authors use HMMs to model rating migration of corporate *bonds*, a factor that affects pricing interest rate margins and subsequently the fair value of corporate *bonds*. Further, Giacomo, Davis, and Crowder (2005) used HMMs to predict default events in a corporate *portfolio*. Under this approach, the hidden states of the postulated HMMs reflect the state of the economy, which can switch between expansion and recession periods (high risk, normal risk), while the emission distributions of each state are taken as binomial distributions modeling the number of defaults in the studied portfolio at a specific point in time. In Banachewicz, Lucas, and van der Vaart (2008), the aforementioned model is extended to include exogenous variables (*covariates*), such as interest rates and GDP. More recently, (Ching, Siu, min Li, Li, & Li, 2009) employed an interactive HMM to model corporate *bond* defaults; this model essentially assumes that the relationship between the hidden state of the economy and the evolution of the creditworthiness of the companies in the modeled *portfolio* is *bidirectional*. Finally, the authors of (Ching, Leung, Wu, & Jiang, 2010) proposed a multistream HMM (MHMM) capable of modeling multiple financial sequences under the assumption that all of them are driven by a common hidden sequence reflecting the state of the economy. They utilize this model to analyze default data in a network of *financial sectors*, and derive reliable estimates of credit value-at-risk (VaR) and expected shortfall for *portfolios* of corporate bonds.

In a different vein, more closely related to our work, MW Korolkiewicz (2008) proposed an HMM-based model that uses credit ratings posted by rating agencies to perform prediction of the future behavior of an obligor (default or non-default), where the behavior variable is modeled as the hidden state of a postulated two-state HMM. More recently, Elliott, Siu, and Fung (2014) proposed a double-HMM approach which extends the method of MW Korolkiewicz (2008) by considering as its observed variables both the credit ratings posted by rating agencies and the calculated Altman Z-scores (Altman, 1968) of the examined companies (two streams of information). Even though these approaches are quite closely related to our work, as they are also dealing with *individual corporate credit rating*, there is also a *key* difference that sets them apart from our work: The applicability of the methods in (Elliott et al., 2014; MW Korolkiewicz, 2008), depends on the availability of credit ratings posted by rating agencies, which is hardly the case when dealing with private companies. In contrast, our work does not impose such severely limiting constraints, but offers a bottom-up architecture aiming to obtain reliable corporate credit scores without provision of any prior (expert) information.

3. Methodological background

3.1. Student's-t hidden Markov models

HMMs are increasingly being adopted in a wide spectrum of applications, since they provide a convenient way of modeling observations appearing in a sequential manner and tending to cluster or to alternate between different possible components (subpopulations) (Cappé, Moulines, & Rydén, 2005). The observation emission densities associated with each hidden state of a continuous density HMM (CHMM) must be capable of approximating arbitrarily complex probability density functions. Finite Gaussian mixture models (GMMs) are the most common selection of emission distribution models in the CHMM literature, yielding the so-called Gaussian HMMs (GHMMs) (Rabiner, 1989). The vast popularity of GHMMs stems from the well-known capability of GMMs to successfully approximate unknown random distributions, including distributions with multiple modes, while also providing a simple and computationally efficient maximum-likelihood (ML) model fitting framework, by means of the expectation-maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977). Nevertheless, GMMs do also suffer from a significant drawback concerning their parameters estimation procedure, which is well-known that can be adversely affected by the presence of outliers in the data sets used for the model fitting. Hence, when outliers are present in the available fitting data sets (as it often happens in real-world applications), GMMs tend to require excessively high numbers of mixture components to capture the long tails of the approximated distributions (corresponding to the existing outliers), so as to retain their pattern recognition effectiveness. As a consequence of the induced model size increase, the computational efficiency of the trained models deteriorates significantly, while high requirements are also imposed in the size of the available training data sets, so as to guarantee the dependability of the model fitting procedure.

As a solution for the amelioration of these drawbacks, the Student's-*t* HMM (SHMM) has been proposed in (Chatzis et al., 2009) as a highly tolerant to outliers alternative to GHMMs. SHMM employs finite mixtures of the longer-tailed multivariate Student's-*t* distribution as its emission distribution models. This selection provides a much more robust approach to data modeling, as training observations that are atypical of a mixture component density are given reduced weight in the calculation of its parameters, under a model-inherent, soundly-founded statistical procedure.

Let us consider a Student's-*t* hidden Markov model comprising I states. Let $\{\boldsymbol{y}_t\}_{t=1}^T$ denote a sequence of observed data points modeled using the considered SHMM. Let us also assume for convenience, and without any loss of generality, that all the hidden state densities of the considered SHMM are approximated by Student's-t mixture models with the *same number* of component distributions, J. Then, from the conditional independence property of the hidden Markov chain (Cappé et al., 2005; Rabiner, 1989) it directly follows that the observations emitted from the same hidden state of the SHMM are independent, identically distributed (i.i.d), such that the probability density of the observation $\boldsymbol{y}_t$ given that it is emitted from the ith model state reads

$$p(\boldsymbol{y}_t; \boldsymbol{\Theta}_i) = \sum_{j=1}^{J} c_{ij} t(\boldsymbol{y}_t; \boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij}, \nu_{ij}) \quad (1)$$

where c_{ij}, $\boldsymbol{\mu}_{ij}$, $\boldsymbol{\Sigma}_{ij}$ and ν_{ij} are the mixing proportion, mean, covariance matrix and the degrees of freedom of the jth component density of the hidden distribution of the ith state of the model, respectively, and $\boldsymbol{\Theta}_i = \{c_{ij}, \nu_{ij}, \boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij}\}_{j=1}^{J}$ $(i = 1, \dots, I)$. The probability density function (pdf) of a Student's-t distribution with mean vector $\boldsymbol{\mu}$, covariance matrix $\boldsymbol{\Sigma}$, and $\nu > 0$ degrees of freedom is (Liu & Rubin, 1995)

$$t(\boldsymbol{y}_t; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \frac{\Gamma\left(\frac{\nu+p}{2}\right)|\boldsymbol{\Sigma}|^{-1/2}(\pi\nu)^{-p/2}}{\Gamma(\nu/2)\{1 + d(\boldsymbol{y}_t, \boldsymbol{\mu}; \boldsymbol{\Sigma})/\nu\}^{(\nu+p)/2}} \quad (2)$$

where p is the dimensionality of the observations $\boldsymbol{y}_t$, $d(\boldsymbol{y}_t, \boldsymbol{\mu}; \boldsymbol{\Sigma})$ is the squared Mahalanobis distance between $\boldsymbol{y}_t$, $\boldsymbol{\mu}$ with covariance matrix $\boldsymbol{\Sigma}$, and $\Gamma(s)$ is the Gamma function, $\Gamma(s) = \int_0^\infty e^{-t} z^{s-1} dz$.

3.2. Model training

Training of the SHMM using multiple training sequences (tokens) can be easily conducted by means of the expectation-maximization (EM) algorithm. Let us consider M independent sequences of fitting data. We assume for convenience, that all the sequences have the same length T, i.e. they comprise T data points, without any loss of generality. Let the mth sequence be $\boldsymbol{y}_m = \{\boldsymbol{y}_{mt}\}_{t=1}^{T}$, $m = 1, \dots, M$, where $\boldsymbol{y}_{mt}$ stands for the tth data point of the mth fitting sequence. Then, from (1), we have

$$p(\boldsymbol{y}_{mt}; \boldsymbol{\Theta}_i) = \sum_{j=1}^{J} c_{ij} t(\boldsymbol{y}_{mt}; \boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij}, \nu_{ij}) \quad (3)$$

or, equivalently, using the properties of the Student's-t distribution (c.f., (Chatzis et al., 2009; Peel & McLachlan, 2000)):

$$p(\boldsymbol{y}_{mt}|\{u_{ijmt}\}_{j=1}^{J}; \boldsymbol{\Theta}_i) = \sum_{j=1}^{J} c_{ij} \mathcal{N}(\boldsymbol{y}_{mt}; \boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij}/u_{ijmt}) \quad (4)$$

where u_{ijmt} is a precision scalar corresponding to the observation $\boldsymbol{y}_{mt}$ given it is generated from the jth component density of the ith hidden state distribution

$$u_{ijmt} \sim \mathcal{G}\left(\frac{\nu_{ij}}{2}, \frac{\nu_{ij}}{2}\right) \quad (5)$$

Let us denote as $\boldsymbol{s}_{mt}$ the state indicator vectors of the observed data, with $\boldsymbol{s}_{mt} = (s_{imt})_{i=1}^{I}$, and

$$s_{imt} \triangleq \begin{cases} 1, & \text{if } \boldsymbol{y}_{mt} \text{ is emitted from the } i\text{th model state} \\ 0, & \text{otherwise} \end{cases}$$

Let us also denote as $\boldsymbol{z}_{mt}^{i}$ the state-conditional mixture component indicator vectors of the observed data, such that $\boldsymbol{z}_{mt}^{i} = (z_{jmt}^{i})_{j=1}^{J}$, and, given that $\boldsymbol{y}_{mt}$ is emitted from the ith state $(s_{imt} = 1)$, it holds

$$z_{jmt}^{i} \triangleq \begin{cases} 1, & \text{if } \boldsymbol{y}_{mt} \text{ is generated from the } j\text{th component} \\ & \quad \text{density of the state} \\ 0, & \text{otherwise} \end{cases}$$

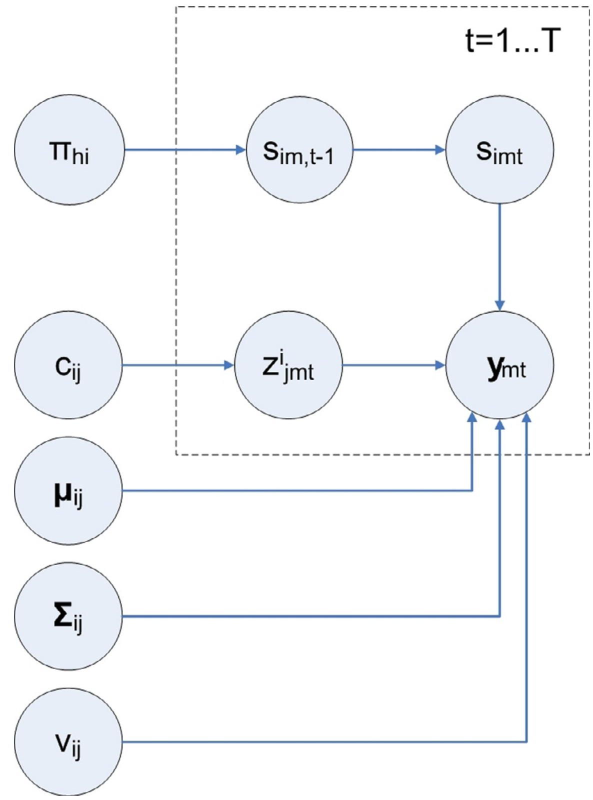


Fig. 1. Directed acyclic graph representing the SHMM (Chatzis et al., 2009). The box (plate) denotes a set of T observation points, $\{\boldsymbol{y}_{mt}\}_{t=1}^{T}$ (of which only a single example for time t is shown explicitly), with their corresponding previous and current state indicators, and their mixture component indicators.

The EM algorithm comprises optimization of the posterior expectation of the complete data log-likelihood of the treated model

$$Q(\boldsymbol{\Psi}; \hat{\boldsymbol{\Psi}}) \triangleq E_{\hat{\boldsymbol{\Psi}}}(\log L_c(\boldsymbol{\Psi})|\boldsymbol{y}) \quad (6)$$

where $\hat{\boldsymbol{\Psi}}$ denotes the obtained estimator of the model parameters vector $\boldsymbol{\Psi} = \{\boldsymbol{\Theta}_i, \pi_i, \pi_{hi}\}_{h,i=1}^{I}$, π_i are the initial state probabilities, and π_{hi} are the state transition probabilities of the Markov chain. For a continuous hidden Markov model, the expression of the complete data log-likelihood reads (Peel & McLachlan, 2000)

$$\log L_c(\boldsymbol{\Psi}) = \sum_{m=1}^{M}\sum_{h=1}^{I}\left[s_{hm1}\log\pi_h + \sum_{i=1}^{I}\sum_{t=1}^{T-1} s_{hmt}s_{im,t+1}\log\pi_{hi}\right] + \sum_{i=1}^{I}\sum_{m=1}^{M}\sum_{t=1}^{T} s_{imt}\log p(\boldsymbol{y}_{mt}^{comp}; \boldsymbol{\Theta}_i) \quad (7)$$

where $\boldsymbol{y}_{mt}^{comp}$ stands for the complete data corresponding to the tth observation of the mth sequence, $\boldsymbol{y}_{mt}$, and $\log p(\boldsymbol{y}_{mt}^{comp}; \boldsymbol{\Theta}_i)$ is the complete data log-likelihood of the emission distribution of the ith hidden state with respect to $\boldsymbol{y}_{mt}$. A graphical illustration (plate diagram) of the considered SHMM can be found in Fig. 1.

To provide a proper complete data configuration for the SHMM, we have to take into account that a closed form solution for log-likelihood optimization of a Student's-t mixture in the form (3) does not exist (Liu & Rubin, 1995; Peel & McLachlan, 2000). However, exploiting the alternative expression (4) and (5) of a Student's-t distribution as a Gaussian distribution with scaled precision, where the scalar is a Gamma distributed latent variable, a tractable optimization framework is obtained. Hence, we let the complete data corresponding to the mth sequence, $\boldsymbol{y}_{m}^{comp}$, comprise the observable data and their corresponding state indicator vectors, state-conditional mixture component indicator vectors, and precision scalars. Then, we have

$$p(\boldsymbol{y}_{mt}^{comp}; \boldsymbol{\Theta}_i) = \prod_{j=1}^{J} \left[c_{ij} p(\boldsymbol{y}_{mt}|u_{ijmt}; \boldsymbol{\Theta}_i) \; p(u_{ijmt}; \boldsymbol{\Theta}_i)\right]^{z_{jmt}^i}$$

which yields (ignoring constant terms)

$$\log p(\boldsymbol{y}_{mt}^{comp}; \boldsymbol{\Theta}_i) = \sum_{j=1}^{J} z_{jmt}^i \Bigg\{ -\log\Gamma\left(\frac{\nu_{ij}}{2}\right) + \frac{\nu_{ij}}{2} \times \left[\log\left(\frac{\nu_{ij}}{2}\right) + \log u_{ijmt} - u_{ijmt}\right] + \log c_{ij} - \frac{u_{ijmt}}{2} d(\boldsymbol{y}_{mt}, \boldsymbol{\mu}_{ij}; \boldsymbol{\Sigma}_{ij}) - \frac{1}{2}\log|\boldsymbol{\Sigma}_{ij}| \Bigg\} \quad (8)$$

The E-step on the (k+1)th iteration of the EM algorithm requires calculation of the quantity $Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k)})$, where $\boldsymbol{\Psi}^{(k)}$ denotes the *current* estimator (obtained by the kth iteration of the EM algorithm) of $\boldsymbol{\Psi}$. Using (7) and (8), we have

$$Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k)}) = \sum_{m=1}^{M} \sum_{h=1}^{I} \left[\gamma_{hm1}^{(k)} \log\pi_h + \sum_{i=1}^{I} \sum_{t=1}^{T-1} \gamma_{himt}^{(k)} \log\pi_{hi} \right] + \sum_{i=1}^{I} \sum_{m=1}^{M} \sum_{t=1}^{T} \gamma_{imt}^{(k)} E_{\boldsymbol{\Psi}^{(k)}} \left(\log p(\boldsymbol{y}_{mt}^{comp}; \boldsymbol{\Theta}_i)|\boldsymbol{y}\right) \quad (9)$$

where $\gamma_{imt}^{(k)}$ denote the kth iteration estimators of the state emission posterior probabilities, defined as

$$\gamma_{imt} \triangleq p(s_{imt} = 1|\boldsymbol{y}) = p(s_{imt} = 1|\boldsymbol{y}_m) \quad (10)$$

$(t = 1, \ldots, T)$, and $\gamma_{himt}^{(k)}$ denote the kth iteration estimators of the state transition posterior probabilities, defined as

$$\gamma_{himt} \triangleq p(s_{im,t+1} = 1, s_{hmt} = 1|\boldsymbol{y}) \quad (11)$$

$(t = 1, \ldots, T-1)$ for $m = 1, \ldots, M,$ $h, i = 1, \ldots, I$. Therefore, the E-step of the algorithm comprises computation of the estimates $\gamma_{imt}^{(k)}$ and $\gamma_{himt}^{(k)}$, and of the expectation $E_{\boldsymbol{\Psi}^{(k)}}\left(\log p(\boldsymbol{y}_{mt}^{comp}; \boldsymbol{\Theta}_i)|\boldsymbol{y}\right)$. Let us begin with the updates $\gamma_{imt}^{(k)}$ and $\gamma_{himt}^{(k)}$. These quantities can be obtained utilizing the forward-backward algorithm. It holds (Cappé et al., 2005; Rabiner, 1989)

$$\gamma_{himt}^{(k)} = \frac{a_{hmt}^{(k)} \pi_{hi}^{(k)} p(\boldsymbol{y}_{m,t+1}; \boldsymbol{\Theta}_i^{(k)}) b_{im,t+1}^{(k)}}{\sum_{\upsilon=1}^{I} \sum_{\phi=1}^{I} a_{\upsilon mt}^{(k)} \pi_{\upsilon\phi}^{(k)} p(\boldsymbol{y}_{m,t+1}; \boldsymbol{\Theta}_\phi^{(k)}) b_{\phi m,t+1}^{(k)}} \quad (12)$$

and

$$\gamma_{imt}^{(k)} = \frac{a_{imt}^{(k)} b_{imt}^{(k)}}{\sum_{h=1}^{I} a_{hmt}^{(k)} b_{hmt}^{(k)}} \quad (13)$$

where

$$a_{im1}^{(k)} = \pi_i^{(k)} p(\boldsymbol{y}_{m1}; \boldsymbol{\Theta}_i^{(k)}) \quad (14)$$

$$a_{im,t+1}^{(k)} = p(\boldsymbol{y}_{m,t+1}; \boldsymbol{\Theta}_i^{(k)}) \sum_{h=1}^{I} a_{hmt}^{(k)} \pi_{hi}^{(k)} \; (t = 1, \ldots, T-1) \quad (15)$$

$$b_{hmT}^{(k)} = 1 \quad (16)$$

$$b_{hmt}^{(k)} = \sum_{i=1}^{I} \pi_{hi}^{(k)} p(\boldsymbol{y}_{m,t+1}; \boldsymbol{\Theta}_i^{(k)}) b_{im,t+1} \; (t = T-1, \ldots, 1) \quad (17)$$

and the expression of $p(\boldsymbol{y}_{mt}; \boldsymbol{\Theta}_i)$ is given by (3). Concerning the term $E_{\boldsymbol{\Psi}^{(k)}}\left(\log p(\boldsymbol{y}_{mt}^{comp}; \boldsymbol{\Theta}_i)|\boldsymbol{y}\right)$, it can be shown that its estimation reduces to computation of the conditional posteriors of mixture component membership

$$\xi_{ijmt}^{(k)} \triangleq E_{\boldsymbol{\Psi}^{(k)}}(z_{jmt}^i|\boldsymbol{y}_{mt}, s_{imt} = 1) = \frac{c_{ij}^{(k)} t(\boldsymbol{y}_{mt}; \boldsymbol{\mu}_{ij}^{(k)}, \boldsymbol{\Sigma}_{ij}^{(k)}, \nu_{ij}^{(k)})}{\sum_{h=1}^{J} c_{ih}^{(k)} t(\boldsymbol{y}_{mt}; \boldsymbol{\mu}_{ih}^{(k)}, \boldsymbol{\Sigma}_{ih}^{(k)}, \nu_{ih}^{(k)})} \quad (18)$$

and of the posterior expectations of the precision scalars u_{ijmt}

$$u_{ijmt}^{(k)} \triangleq E_{\boldsymbol{\Psi}^{(k)}}\left(u_{ijmt}|\boldsymbol{y}_{mt}\right) = \frac{\nu_{ij}^{(k)} + p}{\nu_{ij}^{(k)} + d(\boldsymbol{y}_{mt}, \boldsymbol{\mu}_{ij}^{(k)}; \boldsymbol{\Sigma}_{ij}^{(k)})} \quad (19)$$

Finally, the M-step of the algorithm is effected by performing the computations

$$\pi_i^{(k+1)} = \frac{1}{M} \sum_{m=1}^{M} \gamma_{im1}^{(k)} \quad (20)$$

$$\pi_{hi}^{(k+1)} = \frac{\sum_{m=1}^{M} \sum_{t=1}^{T-1} \gamma_{himt}^{(k)}}{\sum_{m=1}^{M} \sum_{t=1}^{T} \gamma_{hmt}^{(k)}} \quad (21)$$

$$c_{ij}^{(k+1)} = \sum_{m=1}^{M} \sum_{t=1}^{T} r_{ijmt}^{(k)} / \sum_{m=1}^{M} \sum_{t=1}^{T} \gamma_{imt}^{(k)} \quad (22)$$

$$\boldsymbol{\mu}_{ij}^{(k+1)} = \frac{\sum_{m=1}^{M} \sum_{t=1}^{T} r_{ijmt}^{(k)} u_{ijmt}^{(k)} \boldsymbol{y}_{mt}}{\sum_{m=1}^{M} \sum_{t=1}^{T} r_{ijmt}^{(k)} u_{ijmt}^{(k)}} \quad (23)$$

$$\boldsymbol{\Sigma}_{ij}^{(k+1)} = \sum_{m=1}^{M} \sum_{t=1}^{T} r_{ijmt}^{(k)} u_{ijmt}^{(k)} (\boldsymbol{y}_{mt} - \boldsymbol{\mu}_{ij}^{(k+1)})(\boldsymbol{y}_{mt} - \boldsymbol{\mu}_{ij}^{(k+1)})^T \times \left[\sum_{m=1}^{M} \sum_{t=1}^{T} r_{ijmt}^{(k)}\right]^{-1} \quad (24)$$

and solving the equation

$$1 - \psi\left(\frac{\nu_{ij}}{2}\right) + \log\left(\frac{\nu_{ij}}{2}\right) + \psi\left(\frac{\nu_{ij}^{(k)} + p}{2}\right) - \log\left(\frac{\nu_{ij}^{(k)} + p}{2}\right) + \frac{1}{\sum_{m=1}^{M} \sum_{t=1}^{T} r_{ijmt}^{(k)}} \sum_{m=1}^{M} \sum_{t=1}^{T} r_{ijmt}^{(k)} \left(\log u_{ijmt}^{(k)} - u_{ijmt}^{(k)}\right) = 0 \quad (25)$$

to obtain the estimates of ν_{ij}, where, $\psi(s)$ is the digamma function, and $r_{ijmt}^{(k)}$ is the joint posterior probability that $\boldsymbol{y}_{mt}$ is generated from the ith state of the model and particularly from its jth component distribution

$$r_{ijmt} \triangleq p(s_{imt} = 1, z_{jmt}^i = 1|\boldsymbol{y}) = \gamma_{imt}\xi_{ijmt} \quad (26)$$

$$r_{ijmt}^{(k)} = \gamma_{imt}^{(k)} \xi_{ijmt}^{(k)} \quad (27)$$

An outline of the EM algorithm for the SHMM is given in Algorithm 1.

3.3. Inference algorithm

Given a trained SHMM, inference using this model consists in calculating the likelihood of a given sequence, and estimating the emitting (hidden) states sequence corresponding to an observed sequence presented to the model. Let us consider an SHMM, trained using the EM algorithm, as described above, with parameters set $\hat{\boldsymbol{\Psi}}$, and an observed sequence $\boldsymbol{y} = \{\boldsymbol{y}_t\}_{t=1}^{T}$. Then, likelihood calculation can be performed by utilizing the forward algorithm.

Algorithm 1 EM Algorithm for the SHMM.

$k := 0$

1. Conduct the forward-backward algorithm to obtain the quantities $a_{imt}^{(k)}$and $b_{imt}^{(k)}$.
2. Effect the E-step by computing the $\gamma_{himt}^{(k)}$, $\gamma_{imt}^{(k)}$,$\xi_{ijmt}^{(k)}$, $r_{ijmt}^{(k)}$, and $u_{ijmt}^{(k)}$, using (12), (13), (18), (27), and (19), respectively.
3. Effect the M-step by computing the $\pi_i^{(k+1)}$, $\pi_{hi}^{(k+1)}$, $c_{ij}^{(k+1)}$, $\boldsymbol{\mu}_{ij}^{(k+1)}$, $\boldsymbol{\Sigma}_{ij}^{(k+1)}$, and $\nu_{ij}^{(k+1)}$, using (20)-(25), respectively.
4. If the EM algorithm converges, **exit**; otherwise increase theiteration counter ($k := k + 1$) and go to 1.

Specifically, following (Chatzis et al., 2009), the likelihood $p(\boldsymbol{y}|\hat{\boldsymbol{\Psi}})$ yields

$$p(\boldsymbol{y}|\hat{\boldsymbol{\Psi}}) = \sum_{i=1}^{I} \hat{a}_{i,T} \tag{28}$$

where $\{\hat{a}_{i,t}\}_{i,t=1}^{I,T}$ are the forward probabilities corresponding to the observed sequence $\boldsymbol{y}$, computed using (14) and (15) and the parameter estimates $\hat{\boldsymbol{\Psi}}$ of the postulated SHMM. On the other hand, the task of estimating the hidden states sequence corresponding to the observed sequence $\boldsymbol{y}$ can be effected by means of the Viterbi algorithm. Following Chatzis et al. (2009), the estimate of the current hidden state at time t, $\hat{s}_t$, yields

$$\hat{s}_t = \arg\max_{1\leq i\leq I} \delta_t(i) \tag{29}$$

where

$$\delta_t(j) = \max_{1\leq i\leq I}\{p(\boldsymbol{x}_t|y_t = j)\pi_{ij}\}\delta_{t-1}(i), \ t > 1 \tag{30}$$

with initialization

$$\delta_1(j) = p(\boldsymbol{x}_1|y_1 = j)\pi_j \tag{31}$$

4. Proposed approach

4.1. Research motivation

As we have already discussed, corporate credit risk modeling and prediction is typically based on modeling appropriate financial ratio data. On this basis, our approach is motivated from some key insights regarding the nature of the modeled data: It is well understood that creditworthiness patterns exhibit strong temporal dependencies that are reflected in, and can be extracted from, financial ratio data (Buijink & Jegers, 1986; Yli-Olli & Virtanen, 1989). Indeed, corporate credit ratings are well-known to be largely driven by the hidden state of the business cycle process. As such, using machine learning models capable of robustly capturing temporal dynamics in the modeled data is expected to significantly enhance the discriminatory capacity of a developed corporate risk rating system. In addition, outliers and related artifacts are rather common in financial time-series datasets used for model training (Sori, Hamid, Nassir, & Sori, 2006; Whittington, 1980). Therefore, coming up with a modeling method with training algorithms tolerant to the existence of outliers in the used training data is expected to result in better trained models, with enhanced predictive accuracy.

Under this motivation, in this work we suggest to use SHMMs as the core component of a corporate credit rating system, trained on financial ratio time-series. SHMMs satisfy both our requirements of postulating models capable of capturing temporal dynamics in the modeled data, and using models tolerant to outliers in their training data. This formulation is in *stark* contrast to existing machine learning-based approaches used for corporate credit rating, e.g., approaches based on FNNs and RFs, regression techniques, decision trees, and hazard models, which are not capable of extracting temporal dynamics in the modeled data, and, thus, cannot capture changes in the business cycle that could lead in a significant shift in the behavior of the modeled businesses. Another significant merit of our approach that sets it apart from existing approaches is that our use of SHMMs affords modeling *continuous* measured variables pertaining to financial ratios. Finally, as we have shown in Section 3, model training and inference for SHMMs can be performed using robust, elegant, and computationally efficient algorithms with proved convergence (Chatzis et al., 2009). It allows for increased robustness to outliers in the training data, and poses no substantial computational overheads compared to existing competitors.

4.2. System architecture

As depicted in Fig. 2, the proposed corporate credit scoring system comprises four distinct processing stages, namely: (i) data collection and processing; (ii) SHMM model training for each financial ratio; (iii) model aggregation; and (iv) system calibration. We elaborate on each one of these stages in the remainder of this section.

4.2.1. Data collection and processing

This comprises data collection, data processing and transformation, and data selection and samples creation.

Data collection. Training data collection is a significant procedure for the effectiveness of a machine learning system. In the context of our corporate credit rating system, we have collected information on performing and non-performing entities from the supervisory database of the Central Bank of Greece. These data were aggregated during each year from 2006 to 2012, according to the established regulatory framework. The collected information is related with both SMEs and corporations with loans granted from Greek banks; the adopted definition of a default event in this dataset is in line with the rules of Basel III (Basel Committee on Banking Supervision, 2010). Specifically, a loan is flagged as delinquent if it is either 90 days past due or it gets rated as delinquent based on each banks internal rating rules. In the beginning of the observation periods, all considered obligors are performing. In our data collection procedures, we do not consider special cases of obligors from the financial sector, including banks, insurance, leasing, and factoring companies, due to the very unique nature of their business models, which deviate quite a lot from the business models of commercial companies. Under our proposed framework, each obligor is considered to be categorized as either *good* (i.e., performing) or *bad*(i.e., non-performing). Each company was either good or bad at the end of the observation periods. An obligor is categorized as *good*if *at least one of the following criteri*a is met: (i) the obligor manages to get an upgrade of their rating by their bank (according to the bank's internal rules); (ii) the obligor is not delinquent; and (iii) the obligor receives a good rating based on the internal rating system of their bank.

Data selection and samples creation. To create our used datasets, we randomly select good and bad clients from the available population. Each of these clients is represented using a set of financial ratio time-series, extracted using their available balance sheet data and income (P&L) statements. In each case, the dependent variable in our training datasets is a binary indicator, with the on value indicating a default event (i.e., theobligor is categorized as *bad* at the end of the observation period). Finally, some necessary data cleansing is performed on the available data sheets, to remove entries with missing values.This way, we eventually obtain a dataset comprising 8244 obligors, which mainly include Greek

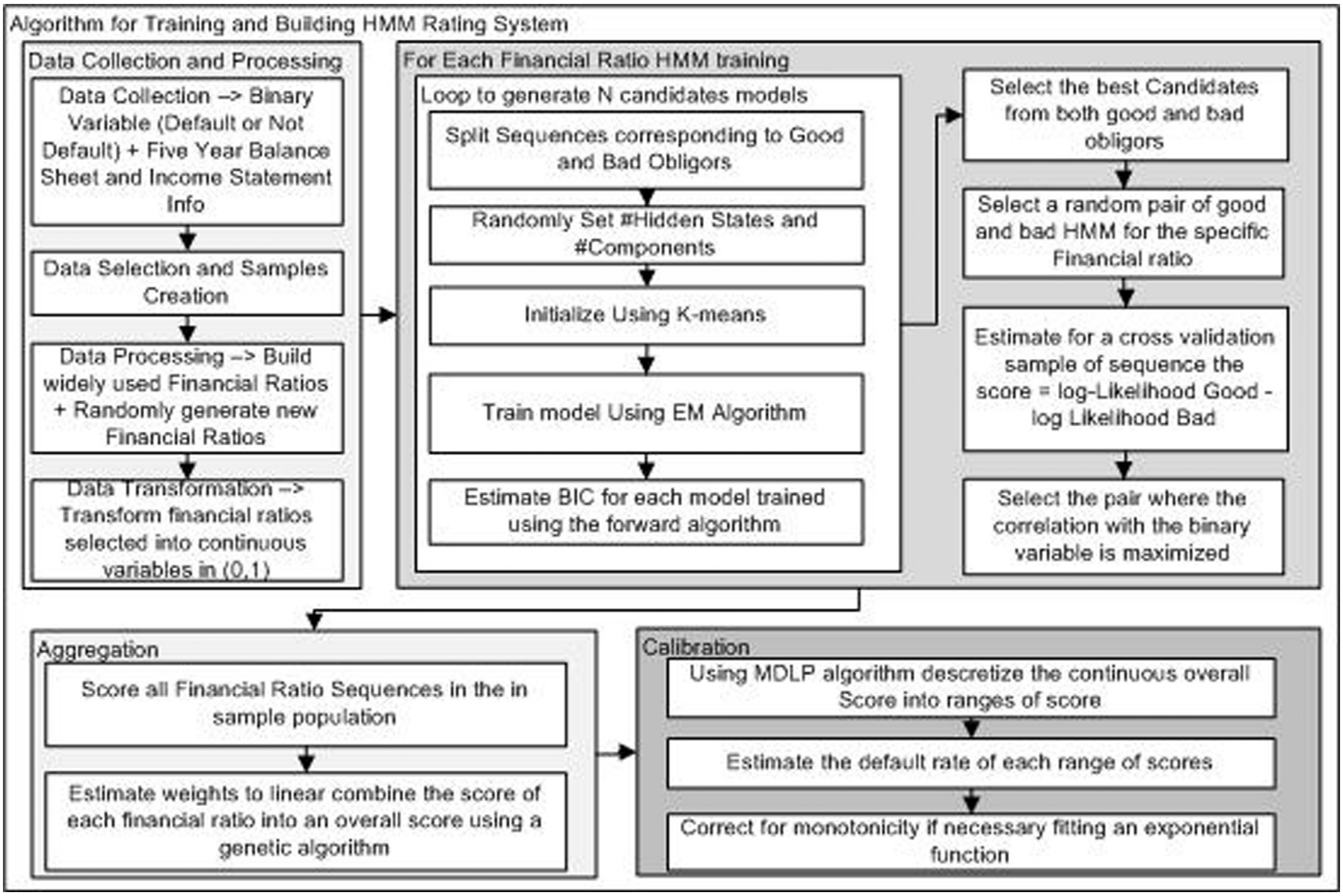


Fig. 2. Proposed system architecture.

Table 1
Dataset split into in-sample, out-of-sample, and out-of-time sets.

Dataset split	Good obligors	Bad obligors	Default rate (%)	Total
In-sample	5513	328	5.62	5841
Out-of-sample	1652	100	5.7	1752
Out-of-time	536	115	17.7	651
Total	7701	543	6.58	8244

Table 2
Distribution of the used data by asset size.

Assets (in Million Euro)	Frequency
5	1837
50	5028
100	707
200	338
500	190
>500	144
Total	8244

SMEs (total assets worth less than 50 Million Euro), as well as some large corporations.

To develop our model, we split the so-obtained dataset into three parts: An*in-sample* dataset, comprising data pertaining to the 70% of the examined companies, obtained over the observation period 2006–2011; an *out-of-sample* dataset, comprising the data pertaining to the rest 30% of the companies for the period 2006-2011; and an *out-of-time* dataset that comprises all the data pertaining to the observation period of year 2012. A summary of the aforementioned split of our dataset is provided in Table 1. In Table 2, we provide a brief summary of the breakout of the used data, showing the numbers of the available samples that pertain to SMEs and large corporations, respectively.

System development, calibration, and analysis is performed using our in-sample dataset. The out-of-sample and out-of-time datasets are in turn used to perform system evaluation under two different scenarios: Evaluation of the generalization capacity of our system across companies, and evaluation of the generalization capacity of our system over time. To allow for reliable estimation of the hyperparameters of our system (i.e., of the number of hidden states, I, of the postulated SHMMs, and the number of components, J, of the entailed Student's-t mixture models), we further split our in-sample dataset into a training set and a validation set: the training sample is used to train the postulated SHMMs pertaining to each financial ratio and each of the two obligor characterizations (*good* or *bad*), while the validation set is used for model selection, i.e. optimal determination of the model hyperparameters (model size).

Data processing and transformation. For each examined corporation, we elect to model a set of well-known and broadly used financial ratio time-series, extracted by exploiting their available balance sheets and income statements. Specifically, the used set of financial ratios comprises the following indices:

- A set of broadly used financial ratios reflecting *liquidity*, including: (i) currentratio (X1); (ii) immediate cash ratio (X2); (iii) working capital (X3); (iv) total employed capital (X4).
- A set of financial ratios that reflect *profitability*, including: (i) return on equity (X5); (ii) return on total employed capital (X6);

(iii) gross profit margin (X7); (iv) operating profit margin (X8); (v) net profit margin (X9).

- A set of financial ratios that reflect *capital structure,* including: (i) fixed assets coverage ratio (X10); (ii) leverage ratio (X11); (iii) interest coverage (X12); (iv) equity over employed capital (X13).
- A set of financial ratios that reflect *activity*, including: (i) receivables turnover ratio (X14); (ii) trade creditors to purchases ratio (X15); (iii) inventories turnover ratio (X16); (iv) employed capital turnover ratio (X17); (v) equity turnover ratio (X18).

In addition to these standard financial ratios, we have also experimented with various transformations of these ratios, with the aim to obtain more representative financial times-series to train our credit rating models with. Specifically, for this purpose, we subsequently followed three distinct procedures:

(i) We applied a series of simple transformations on the original time-series, including *square, cube power, log, sin,1/(1+x),* and *inverse*.

(ii) Subsequently, we computed the year-over-year percentage changes of the considered time-series.

(iii) Finally, in an effort to obtain more robust input variables to train our models upon, we generated 50,000 random (derivative) financial ratios based on the available datasets.For this purpose, we followed an iterative procedure that consists in randomly selecting 4 original items of the original balance sheets and income statements, say *a, b, c*, and *d*, and computing a derivative ratio of the form $(a \pm b)/(c \pm d)$.

This process led to a set of almost 2000 predictor variables (distinct time-series) as potential candidates for our modeling procedures. The so-obtained set of time-series was narrowed down in three consecutive stages: On the first stage, we kept the 200 time-series exhibiting the highest in-sample correlation with the modeled (binary) dependent variable, i.e. the categorization of obligors as *good* or *bad*at the end of the observationperiod.On the second stage, we omitted those of the aforementioned 200 time-series that bear no economic meaning/intuition. Finally, on the third stage, we narrowed down the selected variables (derivative financial ratios) by setting a threshold of at least +/−10% correlation with the *default* flag variable. This way, we eventually retained 8 new *derivative* financial ratio time-series that we use to perform model training in the context of our system, additional to the previously mentioned, commonly used ones. Note that we have transformed the above-mentioned ratio values into the [0, 1] interval, using a simple linear transformation (Kotsiantis, Kanellopoulos, & Pintelas, 2006).

These newly-obtained financial ratio time-series are namely the following: (i) (Operating profit-Interest Expenses)/Sales (X19); (ii) (Short term liabilities + Cost of Sales)/Sales (X20); (iii) (Long term liabilities - Gross profit)/Total Assets (X21); (iv) Bank Loans/Gross profit (X22); (v) (Gross profit + Equity)/Total Liabilities (X23); (vi) Current Assets/Sales (X24); (vii) Borrowed funds/Turnover (X25); and (viii) the *1/(1+x)*transform of the interest coverageratio (X26). Note that the first seven derivative financial ratio time-series mentioned above essentially reflect *capital structure,*thus bearing a clear financial intuition/relevance.

4.2.2. SHMM model training

As previously discussed, for every examined company and for every modeled financial ratio, we build a time-series comprising values recorded over five consecutive years. This dataset is subsequently used to perform model training. As we have also discussed, each sequence is categorized as *good*or *bad,*depending on the corresponding obligor performances. To effectively model these data in the context of the proposed system, we postulate *two distinct*SHMMs for each financial ratio, one pertaining to obligors categorized as *good,*and one pertaining to obligors categorized as *bad.*Our modeling selection allows for capturing salient temporal patterns and dynamics in our modeled time-series, in an effort to detect shifts in the state of the economy and their correlations with changes in the behavior patterns of the examined companies. As we have discussed, we split our in-sample dataset into one training set and one validation set. The training set comprises 200 sequences for each financial ratio, and for each category of obligors (*good*or *bad*), either large corporations or SMEs. Parameter initialization (before model training using the EM algorithm, described in Section 3) was performed using the segmental K-means algorithm described in (Rabiner, 1989); the degrees of freedom ν of the Student's-*t* distributions are initialized at $\nu = 1$.

Table 3
Average optimal size of trained SHMMs by obligor category.

Category	*Good*	*Bad*
# Mixture components	4.00	4.06
# States	5.06	4.09

Turning to model (size) selection, in each case our selections are made between models comprising 2–6 mixture component Student's-*t* distributions, and 2–6 hidden states. To perform model selection, at first we utilize the popular Bayesian information criterion (BIC) (Zucchini & MacDonald, 2009). BIC is widely used for selecting proper model size in the HMM literature, by appropriately penalizing the obtained log-likelihood of the trained model with a penalty term the accounts for the number of postulated parameters (i.e., model size), to prevent overfitting. In the context of our system, we utilize BIC so as to retain the 10 highest-ranked possible model configurations (out of a set of 25 initial alternatives). To alleviate the effect of random initialization (of the segmental K-means algorithm) on the obtained results, we repeat training of each considered model multiple (namely, 20) times, and retain the random restart that yielded the best BIC value. In Fig. 3, we show how BIC values change (on average over the modeled assets) with model size in our experiments, both in cases of models pertaining to *good* obligors, and in cases pertaining to *bad* obligors.

Further, for each of the retained model configurations (sizes), we use the forward algorithm to compute the likelihood of each corresponding time-series in our validation set. We perform this procedure with respect to both the corresponding postulated SHMMs pertaining to obligors categorized as *good,*and the corresponding postulated SHMMs pertaining to obligors categorized as *bad,*and compute the *log-ratio* of the two *likelihoods.*Finally, we rank the postulated alternative SHMM configurations on the basis of the correlation of these obtained (*good* to *bad*)*likelihood log-ratios* with the actual ratings of the considered obligors: the higher the value of these log-ratios, the more correlated they are with obligors actually rated as *good,*and the less correlated they are with obligors actually rated as *bad.*

On the basis of this procedure, for each financial ratio we retain the *good/bad*obligor SHMM pair configuration that yields the best ranking among the considered alternatives. In Table 3, we depict the average obtained SHMM size (over all the modeled financial ratios) separately for the models fitted to data from *good*obligors, and for the models fitted to data from *bad*obligors. As we observe, to sufficiently capture the underlying temporal patterns, companies rated as *good*require larger models than companies rated as *bad.*This is a rather intuitive result, since companies rated as *good*are expected to exhibit more heterogeneous patterns than companies eventually defaulting on their debt.

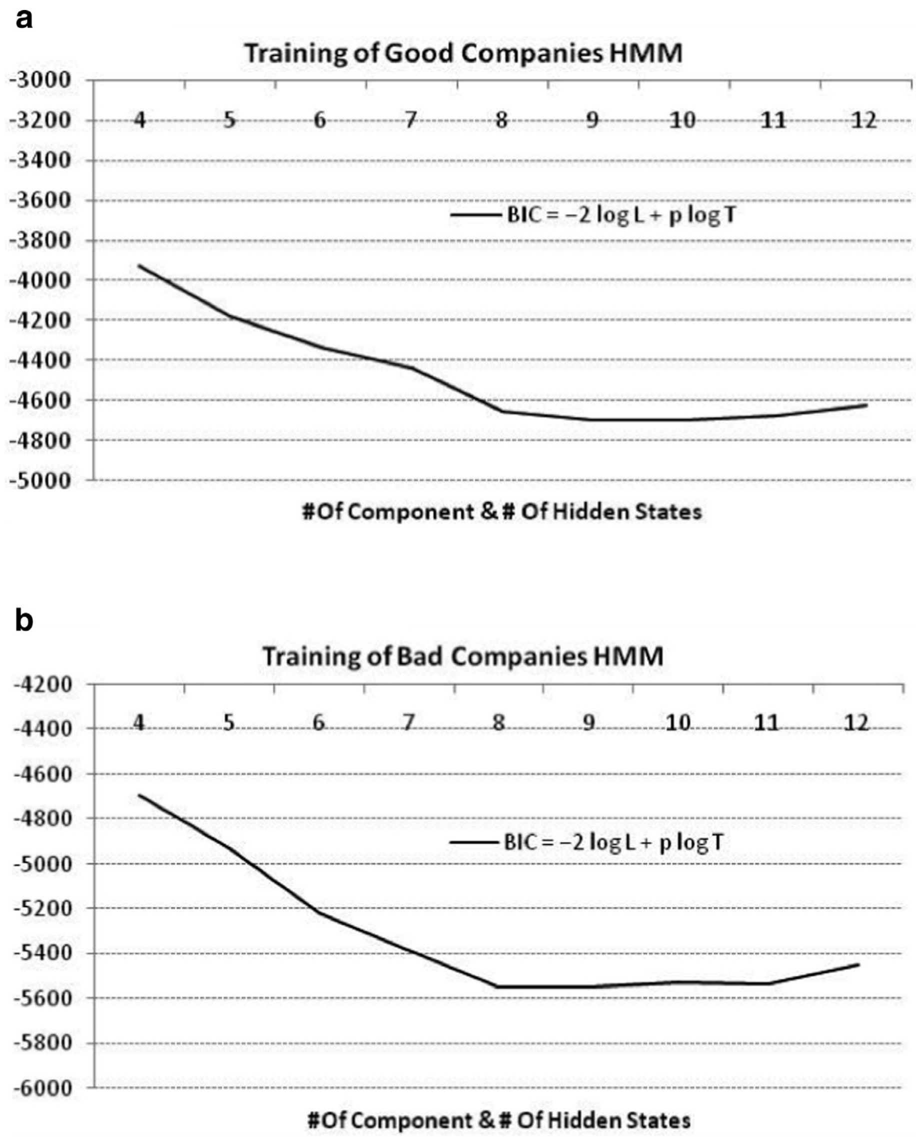


Fig. 3. BIC values as a function of model size.

4.2.3. Model aggregation

After obtaining the component SHMMs of our system, we apply a sample scoring procedure for the entire in-sample dataset. Specifically, for every company in our in-sample dataset, we produce a 26-dimensional vector containing the *likelihood log-ratio scores*pertaining to the two trained SHMMs (*good/bad*) for each modeled financial ratio (for this purpose, the forward algorithm is used as described in Section 3.3). As previously discussed, the *likelihood log-ratio scores*of a modeled company essentially encode how likely our trained models consider the company to end up with a *good*rating at the end of the observation period. Apparently, as a result of our modeling choices, each pair of postulated SHMMs (modeling a different financial ratio) generates a different *likelihood log-ratio score.*Hence, it is necessary that we come up with an optimal way of combining these scores so as to derive a final predictive score from our model.

For this purpose, we postulate a simple linear score combination model driven by the *likelihood log-ratio scores*generated as described previously. To train this model, we use a genetic algorithm (Deb et al., 2000) that aims to maximize the overall score correlation with the dependent variable (good/bad obligor flag) over the modeled in-sample population. Our selection of the aforementioned genetic algorithm as the optimization method of choice is motivated from its simple black-box nature, and its attractive properties it terms of the obtained rates of convergence to the global optimum of the solved complex optimization problem. We experiment with various mutation rates and numbers of generations, in order to select the optimal genetic algorithm configuration. For completeness sake, in Table 4 we provide the final linear model parameter (weight) values estimated through the used genetic algorithm.

4.2.4. System calibration

Eventually, we utilize the credit score values generation capabilities of our system to obtain a *default probability* prediction mechanism. For this purpose, we apply a credit rating system *calibration* process. Calibration of a credit rating system is a mapping process under which each possible generated score value is allocated an associated probability of default. To perform calibration of our system on the in-sample population, we divide the set of obtained *likelihood log-ratio*scores, generated in the previous stage, into ranges. Each range is associated with a probability of default. Computation of ranges is performed in such a way that ensures maximum intra-rate homogeneity of the obtained probabilities of

Table 4
Financial ratio weights obtained by application of the used genetic algorithm.

Financial ratio	Weight
X1	3.8
X2	6.2
X3	9.2
X4	3.4
X5	1.4
X6	6.3
X7	0
X8	2.7
X9	6.1
X10	1.6
X11	1.7
X12	2.4
X13	3.6
X14	1.6
X15	6.5
X16	0
X17	0
X18	1.9
X19	4.9
X20	3.3
X21	3.1
X22	6.6
X23	3.1
X24	5.6
X25	3.1
X26	12.2

default, and maximum inter-range heterogeneity. To achieve this, we use a well known discretization algorithm, namely MDLP; our adopted algorithm follows the minimum description length (MDL) estimation principle (Fayyad & Irani, 1993), which optimizes continuous variable ranges based on a class entropy criterion. Finally, we correct for monotonicity (if needed) by fitting an exponential function.

5. Experimental evaluation

To provide strong *empirical evidence* regarding the merit of our approach, and its utility for expert professionals in the banking sector, we here perform extensive experimental evaluations using real-world data in all cases. We report the performance results obtained from the experimental evaluation of our method, both in terms of out-of-sample performance, and in terms of out-of-time performance.

To obtain some comparative results, apart from our method we also evaluate a set of established benchmark models in the field of corporate credit rating, namely CHAID, LDA, logistic regression, SVMs, RFs, and FNNs. Since the considered benchmark approaches are not capable of modeling time-series data, we opt to retain from (the corresponding time-series of) each financial ratio only those of the five constituent observed variables that do not exhibit strong intercorrelations. Specifically, to perform this procedure, we first compute the Pearson correlation matrix of the available 5-year data (of each financial ratio). On this basis, we exclude the time point variables exhibiting more than 60% absolute correlation with (some) other time point variables, to avoid multicollinearity. Eventually, for each pair of correlated variables, we retain the one that exhibits higher correlation with the dependent variable and drop the other one.

We implemented our method in Microsoft Excel Visual Basic (VBA). We also used the SolveXL add-in of Microsoft Excel to perform genetic algorithm-based optimization. We implemented the MDLP algorithm based on the Discretization package of R.

The remainder of this section is organized as follows: In Section 5.1, we describe the details of our implementation of the considered benchmark approaches (evaluated in parallel to our method). In Section 5.2, we provide an analytical account of our experimental results, and discuss how performance of our method compares to the competition.

5.1. Benchmark models implementation

5.1.1. CHAID

CHAID is well-established algorithm for building decision trees (GV, 1978). Similar to other decision trees algorithms, CHAID allows for simplicity and intuitive visualizations of the obtained results. In addition, the non-parametric nature of CHAID allows for increased flexibility compared to other regression models. Nevertheless, these simplicity advantages come at the cost of significant overfitting proneness due to the entailed discretization of the observed time-series. In our experiments, variable discretization is performed by utilizing 10-bin histograms. We implemented CHAID using the XLSTAT package for VBA.

5.1.2. LDA

LDA is broadly used for credit scoring. For instance, the popular Z-Score algorithm of Altman (1968) is based on LDA. In essence, LDA is used to build binary classification models, predicting whether an examined company will go bankrupt or not. LDA is based on two main assumptions: (i) that the modeled independent variables are normally distributed; and (ii) that the two groups of modeled obligors (*good*and *bad*) exhibit homoscedasticity. As we previously discussed though, these assumptions are hardly plausible in real-world financial time-series. We implemented this approach in R, using the MASS R package.

5.1.3. Logistic regression

Logistic regression is very often used by financial institutions for building credit scoring models due to its parsimonious structure. Similar to LDA, it is used to estimate the non-linear relationship between the modeled continuous independent variables and a categorical/binary dependent variable (in our case, *good* or *bad*obligors). In our implementation, model training is performed using maximum-likelihood estimation. To perform optimization in the context of the M-step of the algorithm, we resort to the Newton-Raphson iterative optimization method. We implemented logistic regression in VBA, using the XLSTAT add-in.

5.1.4. SVMs

SVMs are one of the most popular types of non-linear, large-margin binary classifiers, estimating a separating hyperplane that achieves maximum separability between the data of the modeled two classes (Vapnik, 1998). In our study, we evaluate *soft-margin* SVM classifiers using linear, radial basis function (RBF), polynomial, and sigmoid kernels, and retain the model configuration yielding optimal performance. For the latter purpose, we exploit the available validation set. Similarly, to select the hyperparameters of the evaluated kernels, as well as the cost hyperparameter of the SVM (related to the adopted soft margin), we resort to cross-validation; the candidate values of these hyperparameters are selected based on a *grid-search algorithm* (Vapnik, 1998). We implemented this model in R using the e1071 package; grid-search is a functionality included in the e1071 package (*Tune* routine). The employed cross-validation procedure determines the optimal SVM structure to comprise a linear kernel function with cost hyperparameter equal to 150.

5.1.5. RFs

RFs have recently received considerable attention in various financial research fields (Breiman, 2001). RFs are supervised statistical machine learning methods that combine bootstrap aggregation and random subspace selection to generate or grow trees

Table 5
Discriminatory power results of the evaluated algorithms.

Test	Logistic regression (%)	Neural network (%)	CHAID (%)	LDA (%)	SVM (%)	RandomForests (%)	HMM (%)
GINI							
In-Sample	75.1	75.3	70.6	74.1	71.5	75.1	**79.9**
Out-of-Sample	73.8	74.5	55.2	71.2	65.4	71.5	**80.9**
Out-of-time	78.2	79,5	64.0	75.4	76.0	77.8	**84.3**
K-S							
In-Sample	61.8	62.1	51.7	61.2	59.2	60.2	**67.4**
Out-of-Sample	60.6	63.4	42.3	57.5	54.5	59.2	**68.9**
Out-of-time	65.8	66.7	52.5	63.4	63.1	63.7	**73.3**
Bayesian Error Rate							
In-Sample	24.0	19.0	38.4	25.0	24.7	22.4	**13.9**
Out-of-Sample	17.5	16.4	37.4	18.4	19.7	18.1	**12.8**
Out-of-time	16.6	15.1	32.0	17.8	20.5	16.9	**12.1**

that all together define a forest. In more detail, RFs combine many binary regression decision trees that are selected by bootstrapping samples of the modeled explanatory variables and the corresponding classifier variables. Final prediction is made by averaging the predictions from all the individual trees in cases of regression problems, or using majority voting in cases of classification problems. The final set of random forest variables is selected using a variable importance index, which reflects the "importance" of a variable based on its contribution to classification accuracy. This is estimated by looking at how much prediction error increases when omitting a considered variable. Our implementation of RFs was based on the randomForest package of R. To perform optimal selection of the maximum number of trees in the forest, we perform cross-validation using the available validation set; we select among specifications comprising 20, 50, 100, 200, 500, 600, 700, 800, 900, and 1000 trees. This procedure yields a forest comprising 600 trees. The maximum number of selected variables for each tree is set equal to the 1/3 of the available financial ratios.

5.1.6. FNNs

Typically, credit rating systems employ multilayer perceptron (MLP) FNNs comprising the following layers: the *input* layer, where the explanatory variables are presented to the network, *oneor more* hidden layers comprising sigmoid transfer functions, and an output layer where the predicted values are generated (Bishop, 2006, Chapter 5). To perform model training, we use back-propagation (Bishop, 2006, Chapter 5). We perform cross-validation to select the number of hidden layers and their component hidden neurons, exploiting the available validation set. This procedure selects an MLP with 1 hidden layer and 21 hidden neurons. Training algorithm hyperparameters, including learning rates and momentum values, are also selected by means of cross-validation. Early stopping is employed to avoid overfitting. Our implementation of MLPs was based on the NeuroSolutions toolbox for VBA.

5.2. Comparative results

5.2.1. Discriminatory power results

High discriminatory power is a key requirement for rating systems, and the main evaluation criterion for selecting between alternative rating approaches. To quantitatively measure the performance of a scoring model, researchers and practitioners typically use statistical measures of performance such as the area under the receiver operating characteristic (ROC) curve, the GINI coefficient (accuracy ratio), the Kolmogorov-Smirnoff (K-S) statistic, the Bayesian error rate, Kendall's τ and Somer's D. In this work, we assess the discriminatory power of the evaluated rating systems using the GINI metric, K-S metric, and obtained Bayesian error rates (Basel Committee on Banking Supervision, 2005b). Area under the ROC curve is not used, as it is directly connected with GINI, and essentially captures the same performance characteristics. Similarly, Kendall's τ and Somer's D usually provide similar insights with the aforementioned statistical measures, and, therefore, we decide to omit them from our analyses (Basel Committee on Banking Supervision, 2005b).

In Table 5, we depict the results obtained from the evaluated models. It is evident that the proposed SHMM-based rating system exhibits higher discriminatory power compared to all the considered competitors. More significantly, the obtained performance is more stable and more consistent across all test samples, resulting in lower performance standard deviation. This is an important merit of our approach, since achieving high average performance is as significant for a rating system as it is for it to achieve low performance variance, and thus, higher consistency and better performance guarantees. Another interesting finding stemming from our results is that CHAID performs very poorly in the cases of the out-of-sample and out-of-time datasets. To our perception, this finding is most likely due to overfitting. On the other hand, we observe that FNNs perform slightly better than logistic regression and other machine learning techniques. Note though that the performance superiority of FNNs is not significant enough to counterbalance the advantages of other machine learning approaches, such as LDA and RFs, which offer much better computational complexity, while RFs have also the major advantage of allowing for yielding intuitive visualizations of the results of the inference algorithm.

Regarding the obtained Bayesian error rate, we observe that our results confirm the stability of our approach, since the values of this statistic are similar in all the considered scenarios (in-sample, out-of-sample, and out-of-time), and the classification errors are significantly lower than the considered benchmark models. Finally, we underline that the obtained GINI performance of our model is equal to or greater than 80% in all cases; according to industry benchmarks, SME credit rating systems yielding a GINI index exceeding 80% are considered to possess significantly high (industry-level) discriminatory power. Hence, our approach possesses the significant merit of yielding industry-level predictive performance, which increases its potential attractiveness to real-world financial institutions. A graphical illustration of the evolution of the obtained GINI values is provided in Fig. 4.

5.2.2. Calibration results

Finally, we elaborate on the results obtained from the system calibration procedure described in Section 4.2.4. This procedure yields 9 rating grades. The default rates obtained by the calibrated 9-grade rating system are depicted in Fig. 5 (solid black line).

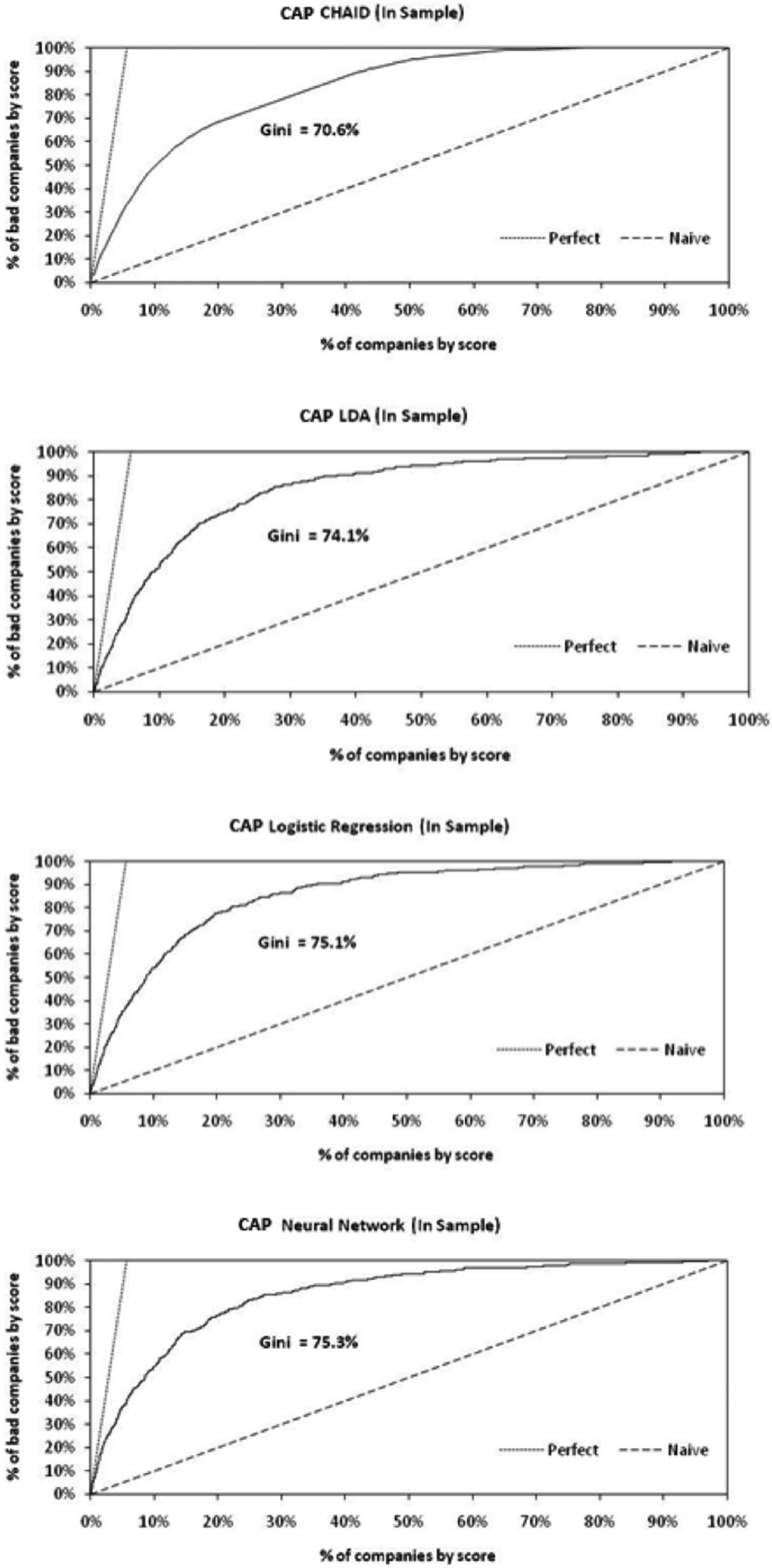
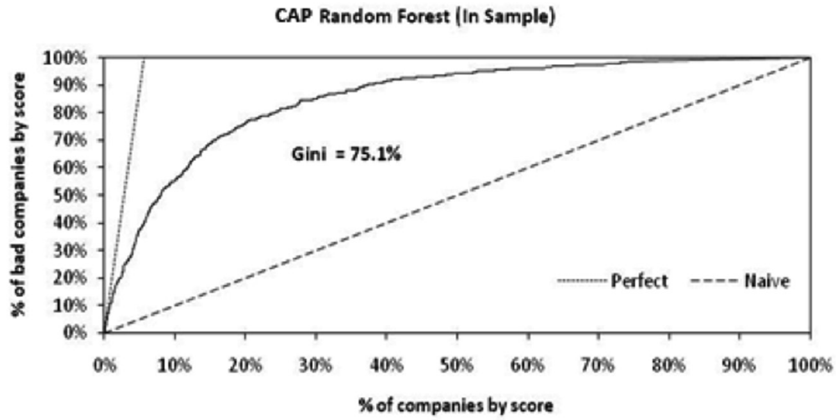
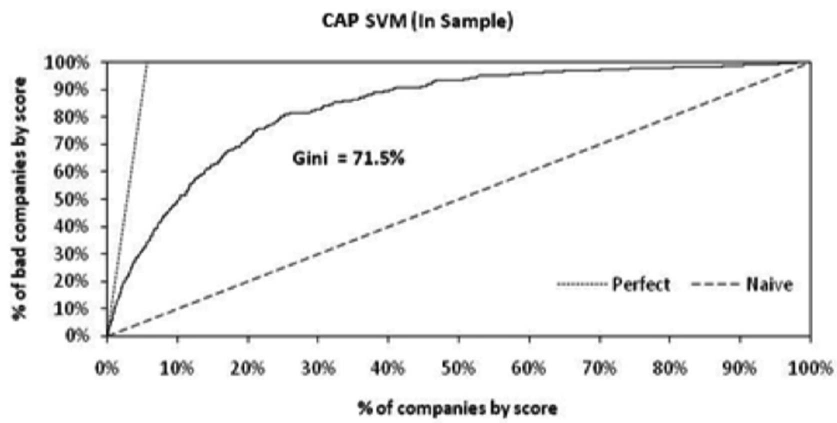
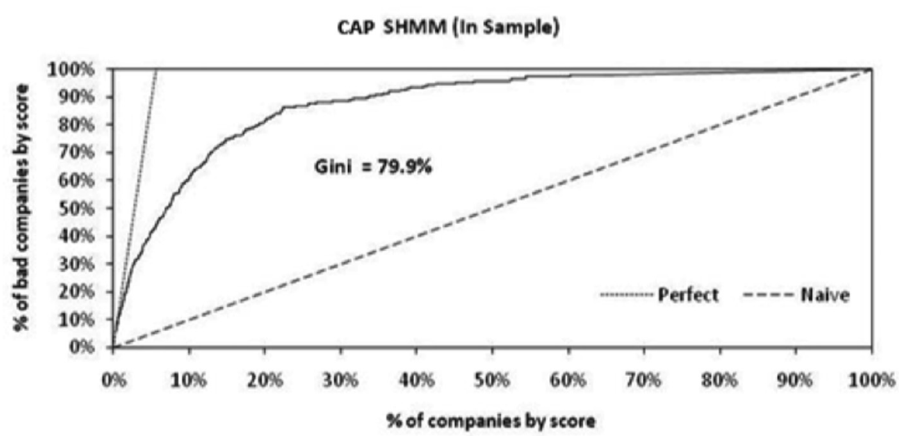


Fig. 4. Analysis of obtained GINI performance values.







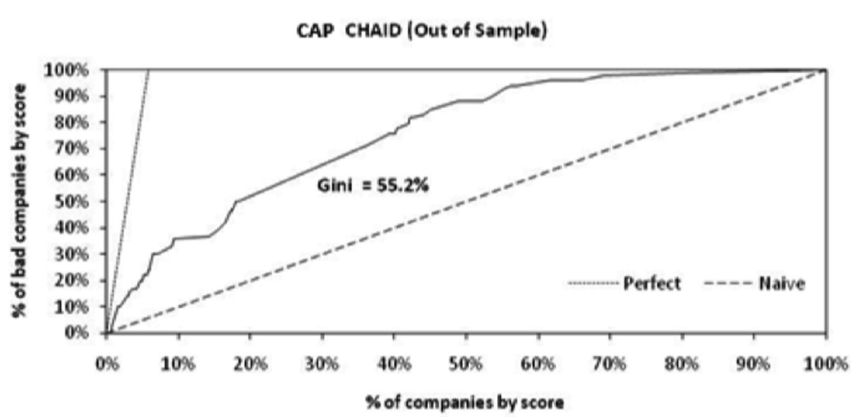


Fig. 4. Continued

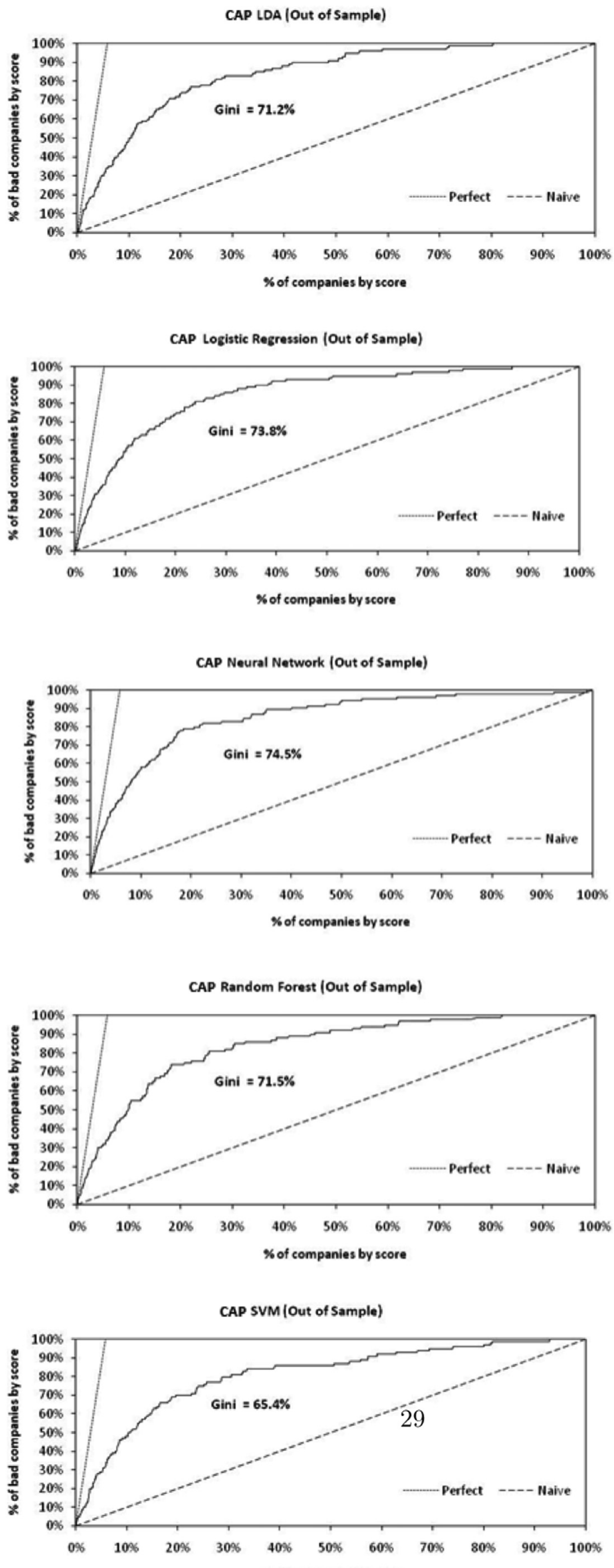
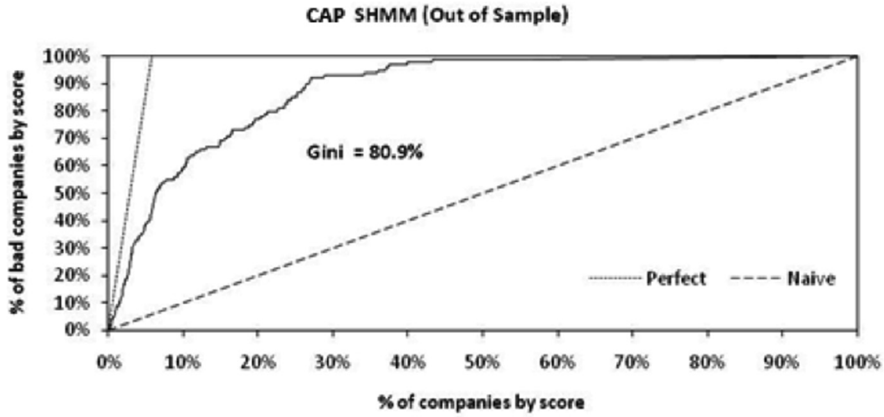
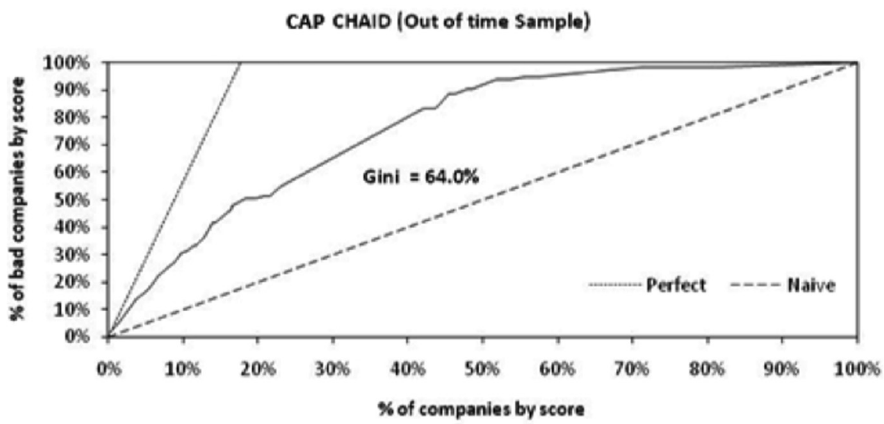
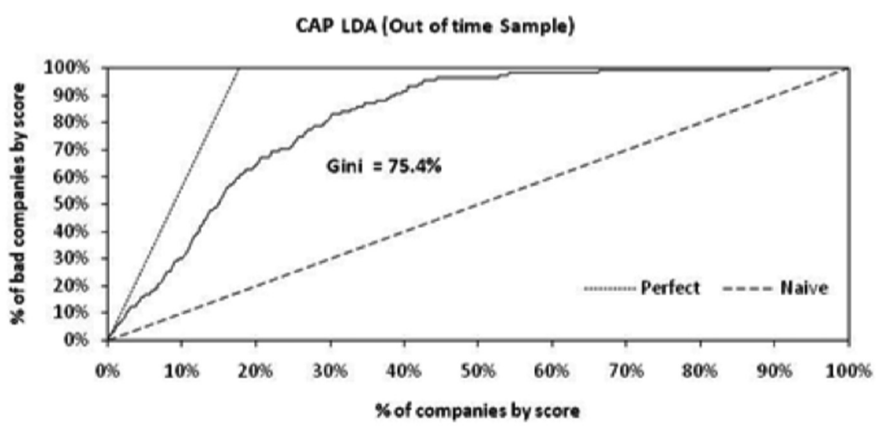


Fig. 4. Continued







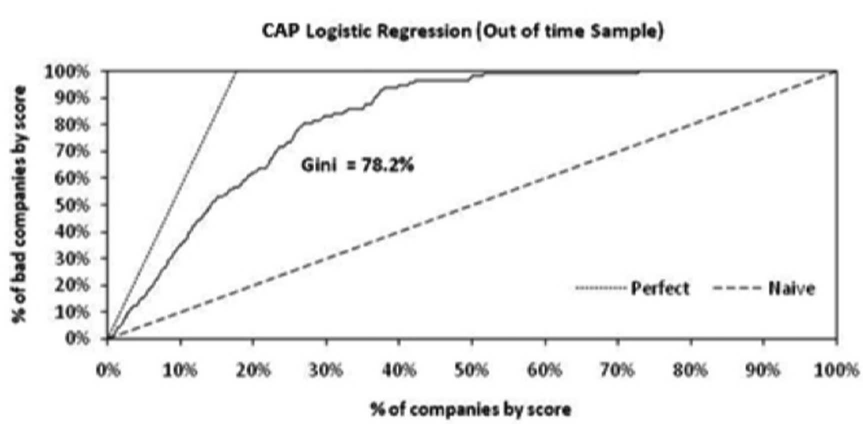


Fig. 4. Continued

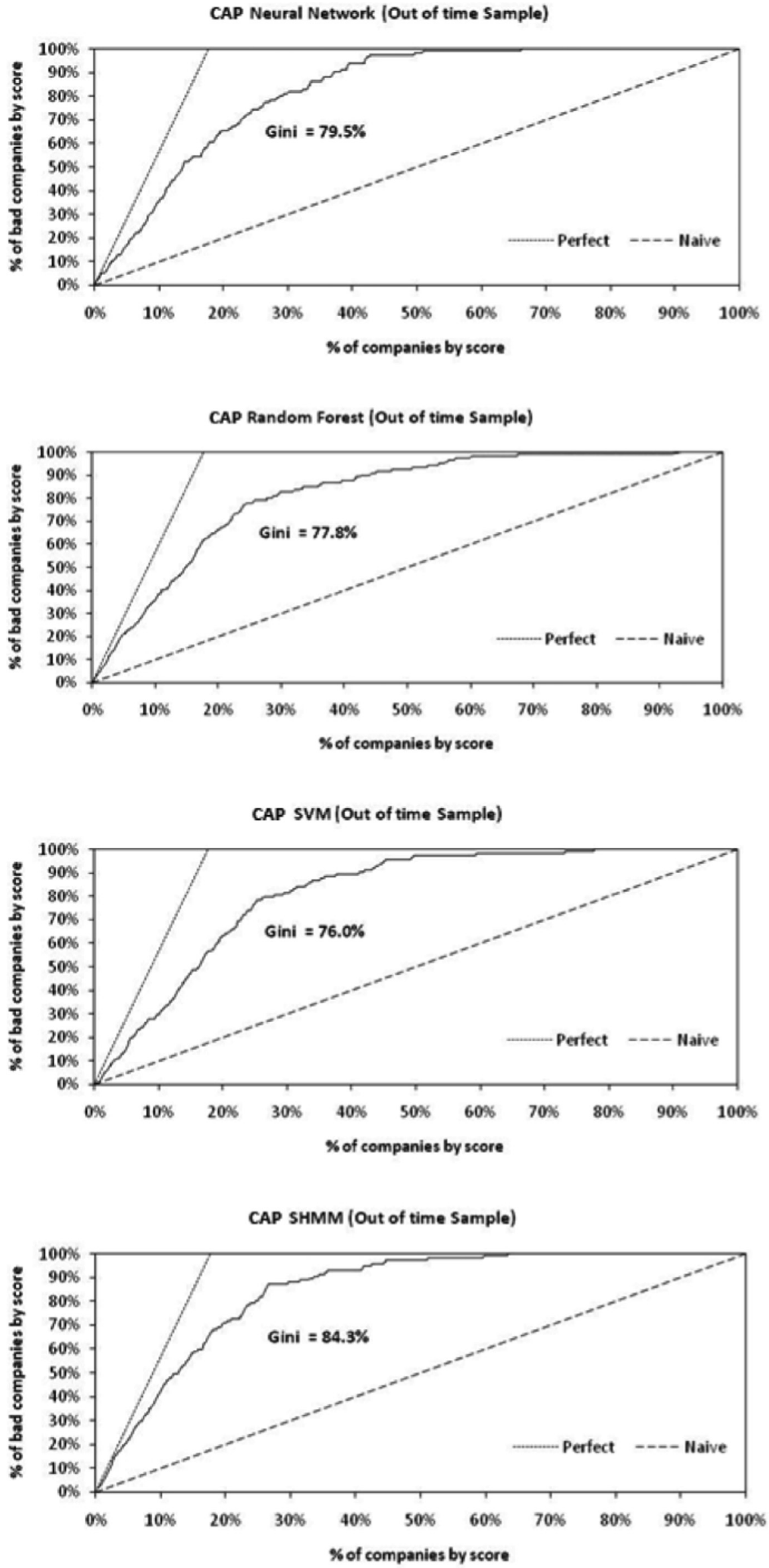


Fig. 4. Continued

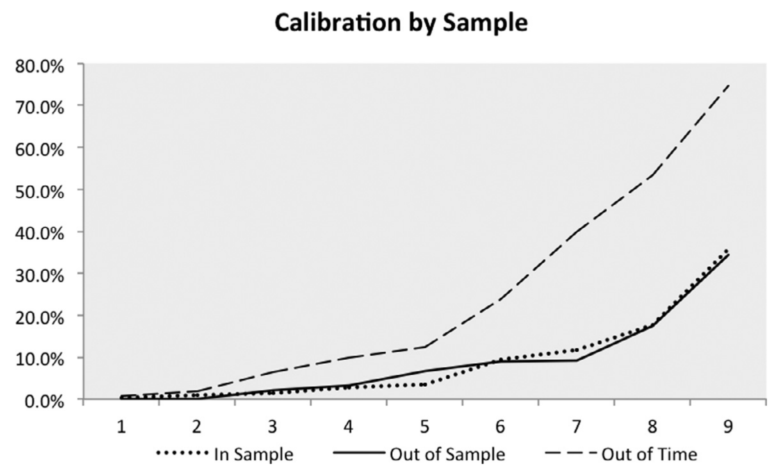


Fig. 5. Calibration results.

Table 6
Calibration results: Chi-squared test outcomes.

	Out-of-sample	Out-of-time
χ^2value (8 DoF)	8.8	288
p-value	>0.10	<0.0001
Null hypothesis	Accept	Reject

In the same figure, we also show the results obtained from applying calibration to the out-of-sample and out-of-time datasets, using the rate ranges determined on the in-sample population. Looking at the out-of-sample results, we observe a rather stable performance in the estimation of the actual default rate that corresponds to the out-of-sample population. This is also verified by a performing a chi-square test (Table 6) to compare the in- and out-of-sample calibrated populations. Therefore, we deduce that our SHMM-based prediction system does not exhibit statistically significant performance differences between the in-sample and out-of-sample datasets.

Further, we perform a similar analysis regarding the out-of-time samples. In this case, we observe quite different a result: Indeed, we observe a deviation of obligors behavior (implied default rate) equal to 5.6% w.r.t. the in-sample dataset, and equal to 17.7% w.r.t. the out-of-sample dataset. We would like to underline that this is not an unexpected system behavior: credit rating systems typically need recalibration in their rating scale when dealing with out-of-time datasets, in order to allow for capturing significant changes in the business environment that cannot be otherwise predicted using the modeled financial ratio time-series. To resolve this issue, one could consider introducing into the fitted models some additional macroeconomic variable as a covariate (e.g., GDP, unemployment rate).

6. Conclusions and future work

In this paper, we proposed a novel credit rating system, leveraging the attractive properties of SHMMs. Our proposed approach is a fully-fledged intelligent system that can be used by financial experts in their corporate credit scoring routines. It addresses all the parts of the modeling pipeline, from financial ratio time-series selection and preprocessing, to selection of appropriate time-series modeling techniques, and information fusion strategies used to obtain the final credit scores.

The main contribution of our work, and its stark difference from previously proposed intelligent systems for corporate credit score rating, consists in the fact that our approach *does not ignore*the heavy-tailed, time-series nature of financial data in generating its predictions. Indeed, the modeling procedures of our system comprise a novel financial time-series modeling scheme based on SHMMs. The utilization of SHMMs allows for capturing intricate temporal dynamics in the modeled data, reflecting the evolution of corporate behavior and risk depending on the latent state of the economy. Furthermore, for the first time in the related literature, we employ an HMM using multivariate Student's-t mixture models as its state emission distributions. This selection allows for us to obtain a model training algorithm with high robustness to *outliers* in the observed datasets, which constitute a common problem in financial time-series data; in addition, it also allows for better capturing correlations between the modeled financial ratios.

To provide strong *empirical evidence*from *real-world application scenarios*regarding the merit of our approach, we performed extensive experimental evaluations of our approach using data from the Central Bank of Greece that pertain to both SMEs and large corporations, recorded over the period 2006–2012. As we showed, our approach consistently outperforms a series of benchmark approaches, both in terms of the obtained GINI coefficients and K-S statistics, and in terms of the obtained predictive variance, which quantifies the model's capacity to retain the high performance levels observed in the in-sample dataset when evaluation is performed using out-of-sample and out-of-time datasets. This performance consistency implies a much stronger generalization capacity compared to the state-of-the-art, which renders our approach much more attractive to researchers and practitioners working in real-world financial institutions, who are mainly interested in the generalization capacity of their systems, rather than in their in-sample performance.

From the *technical* point of view, one aspect that this work did not consider is whether allowing for our model to account for skewness in the observed data could result in yielding even better predictive performances. This is a disadvantage of our approach, which, however, is not addressed either in other existing works. To address this issue, we could consider using mixtures of multivariate skewed-t distributions as the postulated emission distributions (Azzalini & Dalla Valle, 1996; Gupta, 2003), instead of simple multivariate Student's-t distributions. However, the trade-offs between the obtained predictive performance increase and the increased computational costs resulting from such a modeling selection must be thoroughly examined. Further, in this work we have not considered embedding macroeconomic covariates, or covariates pertaining to qualitative information that credit officers may get aware of before it becomes depicted in the balance sheets of the examined companies. However, we underline that none of the risk modeling approaches currently used in the financial world have so far managed to introduce such information into their predictive algorithms.

In addition, we note that in our approach we have postulated univariate component SHMMs, modeling each financial ratio independently of all the others. We opted for this solution to obviate problems arising from multi-colinearities in the modeled data, and to prevent overfitting, which is likely to occur when modeling high-dimensional data. This modeling selection may result in not allowing for the employed dynamic models (SHHMs) to extract salient covariances information, contrary to other approaches, e.g. SVM- and CART-based ones, which are, indeed, capable of extracting such dynamics. Nevertheless, we underline that this weakness does not seem to negatively affect the predictive performance of our approach, which turned out to be superior to the competition in all the considered experimental cases.

From the *managerial*point of view, it is noteworthy that the high discriminatory power of our approach leads to more accurate predictions, hence, more concrete decisions for granting, provisioning, pricing and managing of credit lines. The discriminatory power of a credit rating system refers to its*ex ante*capability to distinguish obligors that are likely to default from those that are not. For instance, an obtained Gini coefficient of 80% (achieved in our experiments) suggests that whenever a good and a bad obligor are rated by our model, the probability of giving a better score to the good obligor is 80%. Furthermore, Stein (2005) provides evidence that the discriminatory power of a credit rating system can make a substantial difference with regard to a banks profitability. In addition, we must underline the observed stability of the discriminatory power of the proposed system across the large variety of considered samples. This merit definitely increases the attractiveness of our system to professional experts, that need an intelligent tool to inform and facilitate their credit risk management procedures, and decreases the need for frequent system retraining; it suffices to merely focus on the recalibration of the rating scale, to capture exogenous economic shocks in the probability of default. We also emphasize that our method is capable of dealing with large datasets of financial ratios, by expanding the training series to capture a*full business cycle.*

Finally, we note that training/calibration of our system takes time similar to existing *advanced* methods, e.g. SVM-based ones, yet up to two times longer compared to the most efficient of the existing methods, that employ *simplistic* assumptions (thus yielding *inferior* predictive performance). This is due to the costs of the EM algorithm, and the entailed forward-backward iterative processes. However, we underline that the system training/calibration procedure is completely *offline*, and is run only *once*. Therefore, the aforementioned extra computational effort linked to system calibration does not undermine the efficacy, the applicability, or the attractiveness of our approach to financial experts and institutions, which will only use the predictive algorithms of our system; the latter are as fast as the most efficient competitors, while also being significantly more accurate, as we show in our experimental evaluations, using real-world data obtained from the Bank of Greece.

References

Altman, E. I. (1968). Financial ratios: Discriminant analysis, and the prediction of corporate bankruptcy. *Journal of Finance, 23*, 589–609.

Avery, R. B. , Calem, P. S. , & Canner, G. B. (2004). Consumer credit scoring: Do situational circumstances matter? *Journal of Banking & Finance, 28*(4), 835–856.

Azzalini, A. , & Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika, 83*(4), 715–726.

Banachewicz, K. , Lucas, A. , & van der Vaart, A. (2008). Modelling portfolio defaults using hidden Markov models with covariates. *The Econometrics Journal, 11*(1), 155–171.

Basel Committee on Banking Supervision (2005a). International convergence of capital measurement and capital standards: A revised framework.

Basel Committee on Banking Supervision (2005b). Studies on the validation of internal rating systems.

Basel Committee on Banking Supervision (2010). Basel III: A global regulatory framework for more resilient banks and banking systems.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.

Breiman, L. (2001). Random forest. *Machine Learning, 45*, 5–32.

Breiman, L. , Friedman, J. , Stone, C. J. , & Olshen, R. (1984). *Classification and regression trees*. CRC press.

Buijink, W. , & Jegers, M. (1986). Cross-sectional distributional properties of financial ratios in belgian manufacturing industries: aggregation effects and persistence over time. *Journal of Business Finance & Accounting, 13*(3), 337–362.

Cappé, O. , Moulines, E. , & Rydén, T. (2005). Inference in hidden Markov models. *Springer Series in Statistics*. New York: Springer-Verlag.

Chatzis, S. P. , Kosmopoulos, D. I. , & Varvarigou, T. A. (2009). Robust sequential data modeling using an outlier tolerant hidden Markov model. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31*(9), 1657–1669.

Chava, S. , & Jarrow, R. A. (2004). Bankruptcy prediction with industry effects. *Review of Finance, 8*(4), 537–569.

Chen, C.-C. , & Li, S.-T. (2014). Credit rating with a monotonicity-constrained support vector machine model. *Expert Systems with Applications, 41*(16), 7235–7247.

Chen, W.-H. , & Shih, J.-Y. (2006). A study of Taiwan's issuer credit rating systems using support vector machines. *Expert Systems with Applications, 30*(3), 427–435.

Ching, W.-K. , Leung, H.-Y. , Wu, Z. , & Jiang, H. (2010). Modeling default risk via a hidden Markov model of multiple sequences. *Frontiers Of Computer Science In China, 4*(2), 187–195.

Ching, W.-K. , Siu, T. , min Li, L. , Li, T. , & Li, W.-K. (2009). Modeling default data via an interactive hidden Markov model. *Computational Economics, 34*(1), 1–19.

Danenas, P. , & Garsva, G. (2015). Selection of support vector machines based classifiers for credit risk domain. *Expert Systems with Applications, 42*(6), 3194–3204.

Deb, K. , Agrawal, S. , Pratap, A. , & Meyarivan, T. (2000). A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II.. In *Lecture notes in computer science: 1917* (pp. 849–858).

Dempster, A. , Laird, N. , & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B, 39*(1), 1–38.

DR, C. (1972). Regression models and life tables (with discussion). *Journal of Royal Statistical Society, Series B, 34*, 187–220.

Elliott, R. J. , Siu, T. , & Fung, E. S. (2014). A double HMM approach to altman Z-scores and credit ratings. *Expert Systems with Applications, 41*(4), 1553–1560.

Fayyad, U. , & Irani, K. (1993). Multi-interval discretization of continuous-valued attributes for classification learning. In *Proceedings of international joint conference on artificial intelligence (IJCAI)* (pp. 1022–1029).

Giacomo, G. , Davis, M. , & Crowder, M. (2005). Analysis of default data using hidden Markov models. *Quantitative Finance, 5*(1), 27–34.

Gupta, A. (2003). Multivariate skew *t*-distribution. *Statistics, 37*(4), 359–363.

GV, K. (1978). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics, 29*(2), 119–127.

Harris, T. (2015). Credit scoring using the clustered support vector machine. *Expert Systems with Applications, 42*(2), 741–750.

Huang, S.-C. (2009). Integrating nonlinear graph based dimensionality reduction schemes with SVMs for credit rating forecasting. *Expert Systems with Applications, 36*(4), 7515–7518.

Huang, S.-C. (2011). Using Gaussian process based kernel classifiers for credit rating forecasting. *Expert Systems with Applications, 38*(7), 8607–8611.

Im, J. , Apley, D. W. , Qi, C. , & Shan, X. (2012). A time-dependent proportional hazards survival model for credit risk analysis. *Journal of the Operational Research Society, 63*(3), 306–321.

Kamstra, M. , Kennedy, P. , & Suan, T. (2001). Combining bond rating forecasts using logit. *Financial Review, 36*(2), 75–96.

Kotsiantis, S. B. , Kanellopoulos, D. , & Pintelas, P. E. (2006). Data preprocessing for supervised leaning. *International Journal Of Computer Science, 1*.

Liu, C. , & Rubin, D. (1995). ML estimation of the *t* distribution using EM and its extensions, ECM and ECME. *Statistica Sinica, 5*(1), 19–39.

Mizen, P. , & Tsoukas, S. (2012). Forecasting US bond default ratings allowing for previous and initial state dependence in an ordered probit model. *International Journal of Forecasting, 28*(1), 273–287.

MW Korolkiewicz, R. E. (2008). A hidden Markov model of credit quality. *Journal of Economic Dynamics and Control, 32*(12), 3807–3819.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research*, 109–131.

Peel, D. , & McLachlan, G. (2000). Robust mixture modeling using the *t* distribution. *Statistics and Computing, 10*, 335–344.

Petr, G. , & Gurný, M. (2013). Comparison of credit scoring models on probability of default estimation for US banks. *Prague Economic Papers, 2*, 163–181.

Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE, 77*, 245–255.

Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business, 74*(1), 101–124.

Sori, Z. M. , Hamid, M. A. A. , Nassir, A. M. , & Sori, S. M. (2006). Some basic properties of financial ratios: Evidence from an emerging capital market. *International Research Journal of Finance and Economics, 2*, 71–87.

Stein, R. (2005). The relationship between default prediction and lending profits: integrating ROC analysis and loan pricing. *Journal of Banking & Finance, 29*, 1213–1236.

Thomas, L. C. , Allen, D. E. , & Morkel-Kingsbury, N. (2002). A hidden Markov chain model for the term structure of bond credit risk spreads. *International Review of Financial Analysis, 11*(3), 311–329.

Vapnik, V. N. (1998). *Statistical learning theory*. New York: Wiley.

Wang, G. , & Ma, J. (2012). A hybrid ensemble approach for enterprise credit risk assessment based on support vector machine. *Expert Systems with Applications, 39*(5), 5325–5331.

Whittington, G. (1980). Some basic properties of accounting ratios. *Journal of Business Finance & Accounting, 7*(2), 219–232.

Yeh, C.-C. , Lin, F. , & Hsu, C.-Y. (2012). A hybrid KMV model, random forests and rough set theory approach for credit rating. *Knowledge-Based Systems, 22*, 166–172.

Yi, B. , & Zhu, J. (2015). Credit scoring with an improved fuzzy support vector machine based on grey incidence analysis. In *Proceedings of IEEE international conference on grey systems and intelligent services (GSIS)* (pp. 173–178).

Yli-Olli, P. , & Virtanen, I. (1989). On the long-term stability and cross-country invariance of financial ratio patterns. *European Journal of Operational Research, 39*(1), 40–53.

Zhao, Z. , Xu, S. , Kang, B. H. , Kabir, M. M. J. , Liu, Y. , & Wasinger, R. (2015). Investigation and improvement of multi-layer perception neural networks for credit scoring. *Expert Systems with Applications, 42*(7), 3508–3516.

Zucchini, W. , & MacDonald, I. L. (2009). *Hidden Markov models for time series: An introduction using R*. CRC Press.