WORKING PAPER SERIES **12-2020**

Measuring the Default Risk of Small Business Loans: Improved Credit Risk Prediction using Deep Learning

by

Yiannis Dendramis, Elias Tzavalis and Aikaterini Cheimarioti

Πατησίων 76, 104 34 Αθήνα. Τηλ.: 210 8203303-5 / Fax: 210 8203499
76, Patission Street, Athens 104 34, Greece.
Tel.: (+30) 210 8203303-5 / Fax: (+30) 210 8203499
E-mail: econ@aueb.gr / www.aueb.gr

Measuring the Default Risk of Small Business Loans: Improved Credit Risk Prediction using Deep Learning*

Yiannis Dendramis   Elias Tzavalis   Aikaterini Cheimarioti

October 20, 2020

Abstract

This paper suggests using a multilayer artificial neural network (ANN) method, known as deep learning ANN, to predict the probability of default (PD) within the survival analysis framework. Deep learning ANN structures consider hidden interconnections among the covariates determining the PD which can lead to prediction gains compared to parametric statistical methods. The application of the ANN method to a large data set of small business loans demonstrates prediction gains for the method relative to the logit and skewed logit models. These gains mainly concern short term prediction horizons and are more apparent for the type I misclassification error of loan default events, which has important implications for bank loans portfolio management. To identify the effects of covariates on the PD by the ANN structure, the paper proposes a bootstrap sampling method obtaining the distribution of changes of the PD over discrete covariate changes, while controlling for possible interactions among the covariates. We find that the covariates with the most important influence on the PD include the delinquent amount of a loan over its total balance, the payments and the balance of the loan over its installment, as well as the delinquency buckets of a loan. The duration of a loan is also found to be an important factor of default risk.

JEL classification: G12, E21, E27, E43

Keywords: business loans, probability of default, deep learning, artificial neural networks, logit, skewed-logit, bootstrap methods, ROC curve, type I and II errors.

*Dendramis: Department of Economics, Athens University of Economics and Business, email: ydendramis@aueb.gr, Tzavalis: Department of Economics, Athens University of Economics and Business, email: etzavalis@aueb.gr, corresponding author, Cheimarioti: Alpha Bank, Greece, Credit Risk Methodologies Division, email: aikaterini.cheimarioti@alpha.gr. The authors would like to thank the participants at the seminars of the Department of Economics of Athens University of Economics Business for useful comments. The views presented in the paper are those of the authors alone and do not reflect official views of their institutions.

1 Introduction

Small business (SB) loans constitute a large proportion of bank portfolio loans, due to their close links to the banking system.[1] Given their high leverage and lack of alternative financing, the SB companies are excess vulnerable to severe recession and financially distressed conditions of the economy. The large numbers of SB companies in the economy in association with their information opacity, weak balances, higher lending rates and trustworthiness of management, as well as corporate structure (corporations or sole proprietors), among others, make credit risk management and, in particular, the prediction of default risk for SB loans (SBLs) a very challenging task (see, e.g.,Dietsch, Petey (2002), Rikkers, Thibeault (2009), Moro, Fink (2013), and Behr et al. (2017)).The limited information and asymmetric problems involved, as well as, the limited sources of data raises important obstacles in developing and testing structural or reduced form models of the probability of default (PD) for SBLs.[2]

This paper contributes into the above literature by proposing the use of an artificial neural network (ANN) structure in predicting the PD of SBLs based on the survival analysis framework. Multilayer ANN structures, also known as deep learning neural networks, can universally approximate complex structures of the covariates driving the PD, known as input variables in the ANN literature. Also, they can uncover relevant information from the data throughout the interconnections among the multiple neurons and hidden layers (hence the synonym "deep learning") that consider. These interconnections, which may be missed in parametric methods, can lead to further gains in the prediction performance of the ANN method. The above features of the ANN method may be proved very useful in predicting default risk, especially under financially distressed conditions where the levels of the NPLs are very high and loan default decisions are affected not only by macroeconomic and credit risk fundamentals, but also on institutional factors whose effects can not be easily measured. These factors include the legislation and bank supervision framework

[1]Note that, according to the European Banking Authority report of 2006 (see EBA/OP/2016/04), in year 2014 the small business and small-medium enterprises of the non-financial business sector were the 99% of the enterprises, employed two in every 3 employees and added 58 cents for every euro of the value added of the GDP in the EU-28 economy.

[2]Note that, since that seminal work of Altman, Sabato (2007) on the financial ratios determining bankruptcy of small business and medium size enterprises, there are few studies modelling and estimating the PD for SBLs.

developed to protect borrowers from default and sometimes can initiate moral hazard attitudes. From a banks perspective, the ANN method can help to minimize possible future losses due to loan default miss-forecasting. Parametric methods (e.g., the logit model), often used in the literature to predict the PD (see Greene (2012), for a survey), may misspecify the true relationships predicting the PD, as they rely on a restrictive set of assumptions.

The above potentials of the ANN method are not without cost. The "black box" nature of the method limits its use in identifying and interpreting covariate effects on PD. For a given dataset there may exist multiple neural network architectures that can generate very similar predictions, thus complicating parameter estimation and identification. In contrast to the parametric methods, the results of an ANN structure are difficult to interpret, and thus the economic evaluation of the covariates relative influence on the PD becomes a challenging task. To this end, in this paper we will adopt a modest approach, in the spirit of that suggested by Gu et al. (2020), recently, to overcome identification issues of the ANN method. Instead of the marginal effects, we focus on identifying covariate effects on the PD based on fitted values of PD between two different percentiles of the distribution of a covariate (e.g., the 10% top and bottom percentiles). This metric can evaluate the impact of any covariate on the PD for a given interpencentile spread of the covariate often used in credit risk assessments. We suggest a bootstrap sampling method to derive the empirical distribution of the generalized effects of a covariate on the PD, while controlling for the effects of other covariates on the PD based on sample information.

We implement the ANN method to a large micro (panel) dataset of Greek SB restructured loans during the period 2014:03 to 2017:06. The percentage of the defaulted loans is very high due to the severe economic recession of this economy, started after the year 2008 global financial crisis. The high rate of loan defaults in our sample helps to more precisely evaluate the prediction performance of the method, as compared to studies in which loan default constitutes a rare event. The rich and disaggregated nature of our data allows us to efficiently evaluate the performance of the ANN method to predict the PD and loan default events, compared to parametric methods. We compare the ANN method to the logit and skewed logit models, frequently used in practice. The skewed logit model, recently suggested in the credit risk literature by Dendramis et al. (2020), is a parsimonious extension of the standard logit model that allows for asymmetric responses of the PD to its covariates.

The estimation of the ANN structure and the two parametric statistical methods considered in our study is based on the following set of covariates: (i) covariates capturing behavioral attitudes and risk sources of the borrowers measured at loan-specific (disaggregate) level, (ii) macroeconomic variables, measured at aggregate level (iii) region and corporate structure qualitative features, and (iv) the duration (age) of a loan account since it was opened. The inclusion of the duration variable into the models enables us to calculate the survival probability of a loan in a future day, conditional on current information and that it has not defaulted until now. The behavioral covariates include the ratio of the delinquent amount of a loan over its total balance, the ratios of the balance and actual payments of the loan over its installment, as well as, the delinquency buckets of a loan measuring the aggregate number of times that a borrower is past due for up to 1, 2 and 3 months, over the history of a loan. These covariates may influence the PD, in addition to macroeconomic ones often used in reduced form PD models in the literature. Studying their effects on the PD can reveal how important are borrowers' risk attitudes in determining the PD, compared to aggregate variables often used in practice. It may be proved very useful to banks and supervisory institutions in monitoring loan quality and managing credit risk. Despite the plethora of empirical studies modelling the PD for mortgages and large companies, to the best of our knowledge, there is a limited number of studies investigating the effects of behavioral covariates such as the above on the PD for SBLs. The ANN approach, throughout the interconnections between its layers and neurons that assumes, may uncover useful information from the loan-specific covariates about the PD across the cross-section and time dimensions of the data.

The results of the paper highlight the usefulness of the ANN method to estimate the PD and predict loan default events. The method compares favorably to the logit and skewed logit models, based on a set of fit and prediction performance metrics applied. In particular, the largest prediction gains of the ANN method relative to the logit models are observed for the *type I* misclassification error of the default event (i.e., a loan is predicted as non-defaulted, while it is actually defaulted) which has important implications for bank loans portfolio management. These gains are more evident over shorter prediction horizons. As was expected, the performance of the ANN method declines as the prediction horizon increases, but still compares favorably to that of the logit models. Regarding the effects of the covariates on the PD, we show that the ANN method provides qualitatively similar results to the parametric methods for most of the covariates considered, but these effects differ in magnitude.

The paper is organized as follows. Section 2 presents the ANN method to predict PD within the survival analysis framework and discuss its estimation method. Section 3 presents the alternative parametric statistical models considered. Section 4 presents the bootstrap sampling algorithm to estimate the generalized covariate effects. Section 5 presents the data and the covariates used in the estimation. Section 6 presents the estimation results; these include in-sample estimates, distribution measures of the covariate effects and an out-of-sample forecasting exercise. Finally, Section 7 concludes the paper.

2 The ANN method of calculating the probability of default

2.1 The ANN structure for survival analysis

Assume discrete time $t = 0, 1, ..., T$ and that, at time t, a loan default event can happen for an individual loan account $i = 1, 2, ..., N$. This event is denoted by a binary choice random variable Y_{it} which takes the value 1 (i.e., $Y_{it} = 1$) if the default event occurs, and zero otherwise (i.e., $Y_{it} = 0$). In survival analysis approach, the probability of default (PD) of a loan i, at time t, denoted as P_{it} is defined as[3]

$$\begin{aligned} P_{it} &= \Pr(Y_{it} = 1 | Y_{is} = 0) \\ &= 1 - \Phi(-\omega_i(x_{it-r}; \theta)), \text{for } s < t,\ i = 1, .., N, \text{ and } t = 1, .., T_i \end{aligned} \tag{1}$$

where function $\Phi(.)$: $\mathbb{R} \rightarrow [0, 1]$ is a cumulative probability function (referred as binary link function) that links the probability of default of a loan account i to an index function of covariates x_{it-r}. This function is denoted as $\omega_i(x_{it-r}; \theta)$, $\omega_i(.)$: $\mathbb{R}^K \rightarrow \mathbb{R}$, where K is the number of elements (covariates) of vector x_{it-r}, and r denotes lag order. The vector θ denotes the set of unknown parameters, and T_i denotes the last observation of a loan account i. The vector $x_{it-r} \equiv [x_{k,it-r}]$, $k = 1, 2, ..., K$, consists of K-covariates which capture

[3]see, e.g., Glennon and Nigro (2005), and Crook and Bellotti (2010), for recent reviews of the survival analysis models.

features of the data often studied in the credit risk analysis. These mainly include the following three groups: (i) covariates reflecting loan-specific features of the data at each point of time t, for all i, referred to as panel data variables (ii) macroeconomic covariates given at aggregate level, which are common across all loan accounts i, and (iii) application covariates which are defined over the cross-section dimension of the data. Note that the application covariates are known at the date that a loan account was opened and they are often given as binary variables (taking 1 and 0 values), reflecting qualitative effects across i. In addition to the above covariates, the index function $\omega_i(.)$ also includes a non linear function of the duration (age) of a loan, denoted as d_{it}. Its inclusion in the index function enables us to calculate the PD of a loan i, or its survival probability, in a future date conditional on a value of d_{it}.

The Artificial Neural Network (ANN) method on predicting the PD (P_{it}) constitutes a semi-parametric approach. This allow us to trace the predictive gains of incorporating non linear covariates interactions of unknown form that are missed by parametric methods. This can be done in several telescopic hidden layers, earning the synonym "deep learning". The ANN approach is considered as the least transparent, least interpretable, and most highly parametrized machine learning tool, but also admitted as the most powerful modelling approach in machine learning. It is theoretically grounded on universal approximation results for any smooth function of predictors (see Hornik et al. (1989) and Cybenko (1989)) and it minimizes any misspecification error of $\Phi(.)$ and/or $\omega_i(x_{it-r};\theta)$ due to an incorrect functional form, by taking into account nonlinear and/or hidden patterns of dependence in the data.

More specifically, the ANN method maps the elements of vector x_{it-r} (the dataset) to an initial layer of multiple units known as neurons. These are then sequentially transferred (mapped) to a new hidden layer of neurons and so on, until the procedure stops to the final layer/node of the ANN which is the predicted default probability, P_{it}, given by relationship (1). Practically, a large enough number of layers and nodes, provides an increased flexibility on modelling the PD by the ANN method. The interconnection among the layers of an ANN structure pass information about features of the data across layers which may lead to a better classification and prediction mechanism (see e.g., Fausett (1994)). In Figure 1, we graphically present the configuration of a multiple neutron/layer ANN to calculate the conditional PD $\Pr(Y_{it}=1|Y_{is}=0)$ for $s<t$.

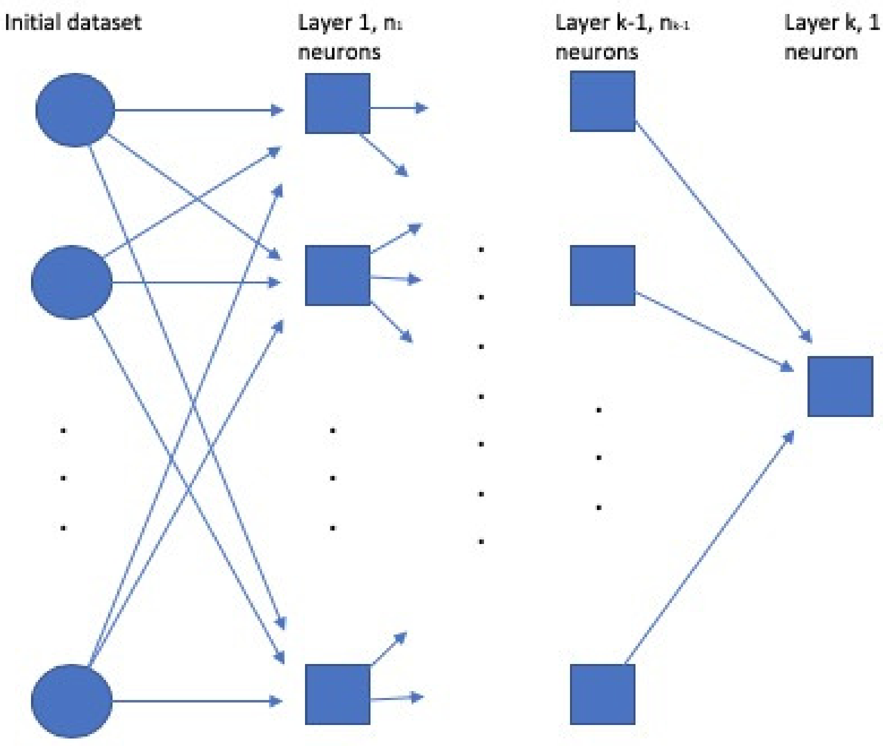

Figure 1: Multiple layers-neurons structure

The connection of the neurons across the sequence of the layers is based on real valued continuous, or discrete, functions known as activation functions. Algebraically, this connection is defined in our panel data set up by the following relationships:

$$neuron_{it-r}^{(v_l,l)} = F^{(v_l,l)}(net_{it-r}^{(v_l,l)};\theta^{(v_l,l)}), \tag{2}$$

$$\text{with} \quad net_{it-r}^{(v_l,l)} = a_0^{(v_l,l)} + \sum_{j=1}^{V_{l-1}} \beta_j^{(v_l,l)} neuron_{it-r}^{(j,l-1)}, \text{ for } i=1,..,N \text{ and } t=1,..,T_i$$

where $l=0,1,2,...,L$ denotes the layers (including the initial one ($l=0$) and the final one ($l=L$)), $v_l=1,2,3,...V_l$ is the number of neurons in each layer l, $net_i^{(v_l,l)}$ is a linear function which aggregates the neurons of the previous $(l-1)^{th}$ layer-step in accordance to their degree of importance, captured by factors $\beta_j^{(v_l,l)}$, and $\theta^{(v_l,l)} = (a_0^{(v_l,l)}, \beta_1^{(v_l,l)}, \beta_2^{(v_l,l)}, ..., \beta_{V_{j-1}}^{(v_l,l)})'$ is the vector of parameters in each layer-step. Note that for the input ($l=1$) and output

$(l = L)$ layers, the functions $net_i^{(v_l,l)}$ are respectively defined as

$$net_{it-r}^{(v_1,1)} = a_0^{(v_1,1)} + \sum_{k=1}^{K} \beta_k^{(v,0)} x_{k,it-r} \text{ and } net_{it-r}^{(1,L)} = a_0^{(v_L,L)} + \sum_{j=1}^{V_{L-1}} \beta_j^{(v_L,L)} neuron_{it-r}^{(j,L-1)}, \qquad (3)$$

since the neurons are replaced with the covariates of the model in the input layer (implying $V_0 = K$) and there exist only one *net* for the output layer, i.e., $net_i^{(1,L)}$. Given $net_i^{(1,L)}$, the conditional PD can be obtained as follows:

$$P_{it} = \Pr(Y_{it} = 1 | Y_{is} = 0) = F^{(1,L)}\left(net_{it-r}^{(1,L)}\right), \text{ for } s < t \qquad (4)$$

where $F^{(1,L)}$ is a logit function, and $F^{(v_l,l)}$, for $l < L$, is the remaining set of activation functions. As it becomes clear from relationships (2)-(4), there is a high degree of flexibility when structuring a neural network. This includes the number of layers l, the number of neurons in each layer V_l, the choice of activation function $F^{(v_l,l)}(.)$, for $l < L$, as well as the aggregation (weights) of information passed through the initial covariates and hidden neutrons to the output function. The universal approximation results of Hornik et al. (1989) and Cybenko (1989) suggests that a single hidden layer is sufficient to approximate any smooth function of predictors. Recent literature suggests that more layers are essential for a better approximation of unknown functional forms, giving rise to the broad acceptance of the deep learning neural networks. Moreover, Barron (1993) provides specific convergence rates for hidden layer feedforward networks with sigmoidal activation functions that can approximate smooth functions. In the literature there are two main strands concerning the choice of the activation function. The sigmoid, or logistic, function $F(z) = \frac{1}{1+e^{-z}}$, $z \in \mathbb{R}$, and the *ReLu* function defined as $ReLu\,(z) = \max\,(z, 0)$.

2.2 Estimation methodology

Let vector θ collect all the parameters of the ANN model $\theta^{(v,l)}$, for all layers l and neutrons v. To estimate this vector, we can use the maximum likelihood (ML) principle. This is equivalent to the minimization of the cross entropy loss function based on the difference between the output Y_{it} and its predicted values, $\widehat{P}_{it}$, frequently used in the machine learning literature.

The log-likelihood function is given as follows:

$$\log L(\theta) = \sum_{i=1}^{N} \log L_i(\theta), \tag{5}$$

where

$$\log L_i(\theta) = \sum_{t=1}^{T_i} Y_{it} \log P_{it} + (1 - Y_{it}) \log(1 - P_{it}), \text{ for } i = 1, .., N \text{ and } t = 1, .., T_i \tag{6}$$

is the log-likelihood function for each loan account i (see **?**). For the numerical optimization of $\log L(\theta)$, we apply brute force optimization using gradient decent methods. Although this increases considerably the computational burden, due to the large number of parameters and the non convexity of the ANN, this approach is considered as more reliable, compared to the commonly used stochastic gradient decent (SGD) approximate method. As frequently noted, the gradient approximation of the objective function performed by the SGD method sacrifices accuracy for the acceleration of the optimization routine.[4]

In general, estimating the parameters of an ANN structure is a challenging task. It requires to choose the optimal number of hidden layers and neurons which fit better the structure into the data. To this end, in our empirical application we will rely on the Bayesian Information Criterion. As is well known, this criterion penalizes heavily models with large number of parameters when this is not accompanied with a large increase in the likelihood value. In addition to the BIC, one can assess the best structure of the ANN approach based on the receiver operating characteristic curve (ROC) (see **?**). This curve can determine the ability of any method to discriminate between default and non-default events, which is crucial in credit risk appraisal (see also Section 5).

3 Parametric survival analysis methods

The comparison of the ANN method to parametric methods predicting the PD is essential in evaluating the benefits of the method in credit-risk management. The most frequently used parametric model of survival analysis is the logit model. This assumes that the bi-

[4]We have also examined the performance of the SGD method, and we have found that its performance is unsatisfactory compared to the brute force optimization method.

nary link function $\Phi(.)$ in relationship (1) is given by the logit distribution, defined as $\Phi(\omega_i(.)) = \frac{1}{1+e^{-\omega_i(.)}}$, where $\omega_i(.)$ is a linear index function of the vector of covariates x_{it-r}, i.e., $\omega_i(x_{it-r};\theta) = \sum_{k=1}^{K} x_{k,it-r}\theta_k$. The estimation of (1) and its statistical inference can be conducted by the ML principle, based on equations (5) and (6). The relative benefits of employing the logit model in predicting the PD compared to an ANN structure stem from the fact that it has a parsimonious parameter structure which can be easily estimated and economically interpreted. The parameter structure of the model implies a much smaller degree of estimation uncertainty, compared to that of the ANN method which usually depends on a very large number of parameters. However, note that the logit model is nested in an ANN structure. It can be obtained from it (see equations (2)-(4)), if the number of hidden layers is set to zero and the activation function is given by the logit distribution

The ANN structure should be also compared with the skewed-logit model suggested recently by Dendramis et al. (2020) in survival analysis. This model is an extension of the logit one. It requires only one more parameter than the logit model to be estimated and it can improve considerably upon the prediction performance of the logit when the response of P_{it} to its covariates has an asymmetric pattern. The binary link function of this model is given by the skewed logit distribution (see, e.g., Burr (1942)), i.e.,

$$\Phi(\omega_i(.);\alpha) = \frac{1}{\left(1+e^{-\omega_i(.)}\right)^{\alpha}} \tag{7}$$

where $a > 0$ is a parameter controlling the asymmetry pattern of P_{it} with respect to its covariates. In particular, a reflects the rate by which P_{it} approaches to 1 (or 0) when responding to the index function of the covariates of the model, i.e., $\omega_i(x_{it-r};\theta)$. If $a < 1$, then $\Phi(.;\alpha)$ approaches to unity with a slower rate than that of the logit distribution ($a = 1$), and the probability density function of $\Phi(.;\alpha)$ is skewed to the right. The inverse happens when $a > 1$. In this case, $\Phi(.;\alpha)$ approaches to unity faster than the standard logit, and the probability density function of $\Phi(.;\alpha)$ is skewed to the left. These results mean that, if $a < 1$ the maximum impact of the covariates of the model occurs at a lower level than $P_{it} = 0.5$, which is assumed by the logit distribution, while, if $a > 1$, at a higher. Note that, for $\alpha = 1$, the skewed logit model reduces to the logit one, assuming a symmetric pattern of P_{it} with respect to $\omega_i(x_{it-r};\theta)$. To estimate the skewed logit model, we can rely on the

ML principle, described before, and the concentrated log-likelihood function for given values of α. The optimal ML estimate of a will be the one that maximizes the concentrated log-likelihood function based a grid search method, over a set of values of a.[5]

4 Generalized covariate effects

An important step of credit risk analysis is the interpretation and the appraisal of the relative effects of discrete, or marginal, changes of covariates x_{kit-r} on the PD, P_{it}, across alternative covariates and/or models. When $\omega_i(x_{it-r};\theta)$ is a linear index function and $\Phi(.)$ is a known parametric distribution (e.g., logit), this can be done by calculating the marginal effect of a covariate x_{kit-r}, $k = 1, 2, ..., K$, on P_{it}, based on the partial derivative of the link function $\Phi(.)$ with respect to $x_{k,it-r}$ and applying the chain rule, i.e.,

$$\frac{\partial P_{it}}{\partial x_{k,it-r}} = \phi\left(-\sum_{j=1}^{K} x_{j,it-r}\theta_j\right)\theta_k \tag{8}$$

where $\phi(.) = \partial\Phi/\partial x_{k,it-r}$ is the partial derivative of the binary link function $\Phi(.)$ with respect to x_{kit-r}. Then, the marginal effects $\partial P_{it}/\partial x_{k,it-r}$ can be easily calculated for given values of covariates $x_{k,it-r}$, for all k, based on formula (8) and the estimates of the vector of parameters θ. In practice, we often calculate $\partial P_{it}/\partial x_{k,it-r}$ at the sample mean values of x_{kit-r} (see, e.g., Greene (2012), for a survey). Since the mean values of x_{kit-r} may not constitute representative values of $x_{k,it-r}$ and/or may not be economically meaningful, Dendramis et al (2020) suggested estimating $\partial P_{it}/\partial x_{k,it-r}$ based on a bootstrap sampling method which relies on sample information of $x_{k,it-r}$ and the remaining covariates of the model $x_{j,it-r}$,for $j \neq k$, on P_{it}, instead of calculating them at their mean values.[6]

In contrast to the parametric statistical models, the calculation of the marginal effects based on an ANN structure raises a number of difficulties, related to the high number of layers and/or neurons that it assumes. This implies an exploding number of parameters

[5] In particular, the concentrated log-likelihood function is given as $\log L(\theta;\alpha)$, where θ denotes the parameters of the index function $\omega_i(x_{it-r};\theta)$. The estimate of α can be obtained by solving the optimization problem: $\widehat{a} = \underset{a\in Q_a}{\arg\max}\left(\log L\left(\widehat{\theta}_\alpha;\alpha\right)\right)$ based on a grid search over values of α, which belong to a reasonable set of α values, Q_a, and $\widehat{\theta}_\alpha$ is the ML estimate for the given value of $\alpha \in Q_a$(see, e.g., Harvey (1990)).

[6] Note that this approach is in the spirit of the generalized impulse response functions (girf) suggested in the literature (see, e.g., Koop et al. (1996)).

to be estimated and raises identification issues, limiting the interpretation of the partial derivatives of P_{it}, as well as our ability to extract statistically identifiable relationships. A multilayer ANN structure implies that there can exist multiple layers and neurons with different weights that can generate very similar predictions of P_{it}, which makes problematic the measurement of reliable marginal effects for a given set of data.

To overcome the above difficulty and take advantage of the benefits of the ANN method in providing accurate predictions, we can evaluate the distinct effects of $x_{k,it-r}$ on P_{it} by calculating changes of P_{it} due to discrete changes in $x_{k,it-r}$, instead of relying on the marginal effects $\partial P_{it}/\partial x_{k,it-r}$. An appropriate such metric for credit risk management would be to calculate changes of P_{it} between two distinct percentiles of the sample distribution of $x_{k,it-r}$ taking also into account the distribution of the other covariates $x_{j,it-r}$, for $j \neq k$. To present this metric more formally, consider the bottom and top 10% percentiles of the sample distribution of a covariate $x_{k,it-r}$ (known as deciles), i.e., $x_{k,it-r}^{(10\%)} = \left\{x_{k,it-r} \leq \Phi_{x_k}^{-1}(10\%)\right\}$ and $x_{k,it-r}^{(90\%)} = \left\{x_{k,it-r} \leq \Phi_{x_k}^{-1}(90\%)\right\}$, where $\Phi_{x_k}^{-1}(.)$ denotes the quantile function of the marginal distribution of $x_{k,it}$, denoted as $\Phi_{x_k}(.)$. Then, define the following sets of the sample points pairs $(i, t-r)$:

$$I_{k,it-r}^{(10\%)} = \left\{(i, t-r) : x_{k,it-r} \in x_{k,it-r}^{(10\%)}\right\} \text{ and } I_{k,it-r}^{(90\%)} = \left\{(i, t-r) : x_{k,it-r} \in x_{k,it-r}^{(90\%)}\right\}. \quad (9)$$

Given $I_{k,it-r}^{(10\%)}$ and $I_{k,it-r}^{(90\%)}$, we can calculate the change of P_{it} between the two deciles of $x_{k,it-r}$, $x_{k,it-r}^{(10\%)}$ and $x_{k,it-r}^{(90\%)}$, for all pairs $(i, t-r)$, based on following metric:

$$\Delta P_{k,it} = P_{k,it}\left(x_{it-r} | (i, t-r) \in I_{kit-r}^{(90\%)}\right) - P_{k,it}\left(x_{it-r} | (i, t-r) \in I_{kit-r}^{(10\%)}\right), \quad (10)$$

where x_{it-r} is the vector of all covariates $x_{k,it-r}$, for $k = 1, .., K$. Apart from the interdecile effects of changes of covariate $x_{k,it-r}$ on P_{it}, equation (10) also takes into account the effects of the remaining covariates $x_{j,it-r}$, for $j \neq k$, on P_{it} which may interact with $x_{k,it-r}$ and influence $\Delta P_{k,it}$. This approach is along the lines of that suggested by Dendramis et al (2020) to calculate generalized marginal covariate effects for parametric PD models, discussed above. Note that the approach can be easily applied to any distinct percentile of $x_{k,it-r}$.

Next, we present the steps of a bootstrap sampling method to calculate the sample distribution of metric $\Delta P_{k,it}$ based on the estimates of an ANN structure, or any of the two of the logit models considered.

Generalized Covariate Effects:

step 1: From the sample distributions of deciles $x_{k,it-r}^{(10\%)}$ and $x_{k,it-r}^{(90\%)}$ draw values of covariate $x_{k,it-r}$ and obtain sets $I_{k,it-r}^{(10\%)}$ and $I_{k,it-r}^{(90\%)}$

step 2: Sample with replacement 1000 values of the whole vector of covariates x_{it-r} for the pairs of points $(i, t-r) \in I_{k,it-r}^{(10\%)}$ and $(i, t-r) \in I_{k,it-r}^{(90\%)}$

step 3: Estimate $\Delta P_{k,it}$ based on the bootstrap values of vector x_{it-r} obtained in step 2 and the estimates of ANN structure. Denote these estimates as $\Delta \hat{P}_{k,it}$

step 4: Compute the empirical distribution of $\Delta \hat{P}_{k,it}$

The distribution of $\Delta \hat{P}_{k,it}$, obtained by the above procedure, can be used to calculate the median and/or different quantiles of $\Delta \hat{P}_{k,it}$ which can be employed to assess the relative effects of a covariate $x_{k,it-r}$ on P_{it} controlling for the effects of the remaining covariates $x_{j,it-r}$,for $j \neq k$.

5 Data on Small Business Loans

To obtain estimates of the PD for the ANN structure and the two logit models presented in the previous sections, we will use a large set of SBL accounts ($N = 11522$), provided by one of the largest systemic banks of Greece.[7] The data have monthly frequency and cover the period from 2014:03 to 2017:06. During this period, the Greek economy was under severe recession and the number of non-performing loans (NPLs) was growing fast. The total number of observations used in our empirical analysis is 268089 (=$T \times N$). Following the Basel II definition of default, we assume that an SB loan is defaulted (i.e., non-performing) when its principal and payments of interest are past due by 90 days, or more. The maturity horizon of the loans varies from 4 to 40 months, while a large proportion of them are referred to sole proprietors (about 60%).

[7]For confidentiality reasons, all identification information (e.g., names, addresses, security numbers were removed by the bank before proving the data, following all security protocols.

Our data set has two interesting features. The first is the high percentage of the defaulted loans (almost 40%), due to the distressed financial conditions held in the economy during the sample interval. This is expected to improve, considerably, the efficiency of the estimation and inference procedures.[8] The second feature is that our data consists of restructured loans increased under the legal procedures introduced by the government (e.g., laws 3869/2010, 3986/2011 and 4224/2013) to deal with the problem of NPLs (see, e.g., Charalambakis et al. (2017)). These procedures favor loan restructuring based on credit ability related criteria, such as borrowers' income, employment status, family size and potential disabilities, inter alia. They are aimed to help borrowers to service their debt, under severe recessions and systemic financial crises (see, e.g., Dergthaler et al (2015), for a survey). [9] For banks, they constitute a prominent policy tool to protect the values of their credit portfolios and avoid weakening their balance sheets. However, for borrowers, they may raise moral hazard incentives. The proposed survival-ANN method may better capture the above features of the data, through the increased flexibility that it offers in modeling hidden relationships across covariates, reflecting complicated economic conditions, such as those observed in our sample.

5.1 Covariates

The covariates used in our study can capture behavioral attitudes of borrowers and the effects of changes in the macroeconomic conditions of the economy on the loan default decisions. Next, we present these covariates in more detail, categorizing them into three main groups, and we discuss their expected effects on the PD based on the theory and empirical evidence reported in the literature.

(i) Application covariates

This group includes the following qualitative dummy variables, taking values 1 if a specific characteristic holds, and 0 otherwise: (*a*) *urban area*, capturing companies in the broad area of Athens and (*b*) *sole proprietorship* (denoted as SP), characterizing companies owned by a single person. We expect the sole proprietorship (SP) covariate to be negatively

[8] See King, Zeng (2001).

[9] Note that a restructured loan is often treated as a new entry by the banks, with a new account opened at the time that the old loan is defaulted and exits the sample.

associated with the PD, due to internal-control mechanisms of the enterprises reducing asymmetric information problems between the bank and the borrower (see, e.g., Fich and Slezak (2008)).

(ii) Loan-specific covariates

In this group we include the loan-specific covariates defined over loan accounts i and time-points t of the sample. These include the ratio of the delinquent amount of the loan over its total balance (denoted as $delinq.amount/balance$), and the ratios of the actual payments of the loan and its balance over its installment (denoted as $payment/installment$ and $balance/installment$, respectively). In these ratios, note that the total balance of a loan includes arrears such as penalties, unpaid installments, extra interest rates etc. The installment of a loan is contracted at its application date. On the other hand, the actual payments of a loan can change over time depending on the ability of the borrowers to fully service their debt. Under the economic conditions that characterize our data, note that banks might receive payments less than the installments in order to help borrowers to service their debt.

Apart from the inability to pay, due to recession and/or liquidity constraints, the above ratios can mirror sources of default risk reflecting strategic decisions of borrowers to default and/or behavioral attitudes, like moral hazard incentives. That is, borrowers with ability to pay and without liquidity problems may strategically not pay back their loans (see, e.g., Dewatripont, Tirole (1994)). Moreover, the ratio $delinq.amount/balance$ is often considered as a measure of consistency of the borrower to pay his/her loan, and thus it can be influenced by moral hazard incentives leading borrowers to delay their loan payments. The effect of the above ratio on the PD is expected to be positive. Another measure of loan payment consistency often used, in practice, is the number of times that a delinquent borrower has a positive amount in buckets 1, 2 and 3, cumulatively. That is, the loan is past due for up to one, two and three months, over its history. This measure, referred to as the number of delinquency buckets (*no.delinq.buckets*), can be considered as an alternative to the $delinq.amount/balance$ one to capturing possible moral hazard incentives on the PD, although a complete characterization of such incentives is hard to be made.

Finally, the ratio $balance/installment$ is expected to have a positive effect on PD, which can be explained by the same reasoning with $delinq.amount/balance$. For the ratio $payment/installment$, one can argue that a negative sign of its on the PD (implying a decrease in the PD) can be interpreted as evidence that the borrowers are committed to meet their obligations, meaning that their actions are not driven by moral hazard.

(iii) Macroeconomic Covariates

This group of covariates consists of the *inflation* and the real gross domestic product (*gdp*) growth rates of the economy. Both series are seasonally adjusted. The quarterly series of the real gdp growth rate is interpolated on monthly frequency. The above macroeconomic variables can capture nominal and real effects of changes in the business cycle conditions. These effects constitutes the two main aggregate sources of default risk in the economy, for all borrowers, and they can influence the ability of the small businesses to pay back their loans. As is well known in the literature (see, e.g., Bergthaler et al. (2015)), credit riskiness of small and medium sized firms exhibit a cyclical pattern following the changes in business cycle conditions, while default rates increase, substantially, during recessions.

Figures 2A graphically presents the series of default rates of the SB loans, over the sample interval 2014:03-2017:06, while Figure 2B presents a graph of these rates conditional on the duration time (defined in months) of the loans. The conditional (average) default rates are aggregated over the cross-section dimension of the data, for all duration intervals presented in the horizontal axis. Inspection of the Figure 2A shows that, until the middle of year 2015, the SBLs defaults exhibited abrupt shifts. Since then, they grown with an approximately constant rate until the end of the sample. As Figure 2B indicates that the highest default rate is reached at shorter loan durations (i.e., up to one year). This is smaller than that observed in other loan markets (e.g., mortgages and consumer loans, see Deng et al. (2000) and Dendramis et al. (2020) and Dendramis et al. (2009)). Almost certainly, this can be attributed to the shorter maturity of the SBLs.

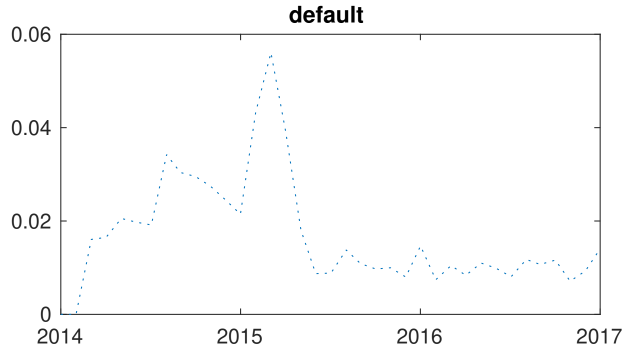

Figure 2A: Default rates over time

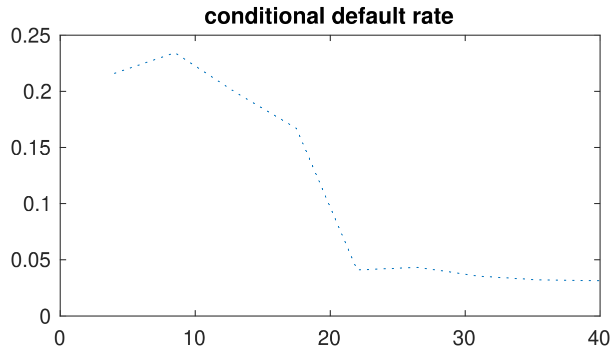

Figure 2B: Default rates conditional on loan duration (age)

In Figure 3, we present the time varying distribution of the behavioral variables *balance/installment*, *payment/installment*, *no.delinq.buckets*, *delinq.amount/balance*, through a fan chart visualization. Inspection of this figure indicates that. for all the above covariates, the cross-sectional dispersion is substantial. For *delinq.amount/balance* and, more evidently, *payment/installment*, the interpencentile spread is larger that than that for *balance/installment*. The covariate *no.delinq.buckets* posses an upward trended volatility, highlighting the increased tendency of borrowers to default over time. The substantial differences in the above interpencentile spreads may help to better identify the covariate effects on $\Delta P_{k,it}$ and, thus, facilitate their interpretation for both the ANN method and the two logit models considered.

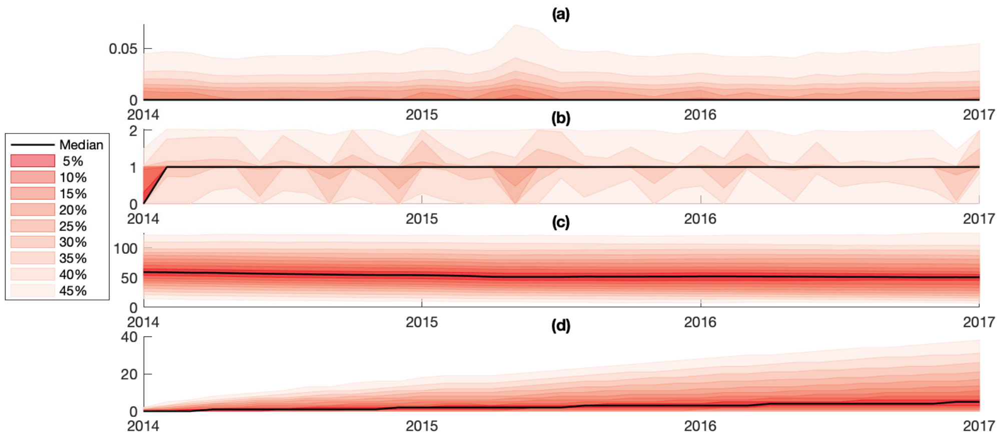

Figure 3: Time varying distribution (fan chart visualisation) of the behavioral covariates: (a) is delinquency over balance, (b) is payment over installment, (c) is balance over installment and (e) is number of delinquent buckets.

6 Estimation results

6.1 In-sample estimates

We present two different sets of estimation results for the ANN method and the two parametric models considered, i.e., the logit and the skewed logit. The first assumes a lag structure of $r = 3$ months for relationship (1), while the second assumes $r = 12$ months. The lag structure $r = 3$ constitutes a natural choice, since loan default occurs when principal and payments of interest are past due by 90 days (i.e., 3-periods back from t).[10] The lag structure $r = 12$ can indicate the ability of all the above methods to provide forecasts of default events over a longer horizon ahead, which is often of interest for bank regulators and/or monitoring authorities. From relationship (1), it can be seen that, at time $t - r$, the lag structures $r = 3$ and $r = 12$ can be also thought of as a shorter and longer prediction horizons.

Since the features of the SB enterprises differ across alternative corporate structures (i.e., corporation vs sole proprietorship), in our analysis for the two logit models we distinguish the effects of the behavioral variables on the PD between the corporation (denoted as $Corp$) and sole proprietorship (denoted as SP) structures, by using appropriately defined dichotomous variables. For the ANN method, this is, in principle, unnecessary. A sufficient

[10]Note $r = 3$ is also the best chosen lag structure based on the Bayesian and Akaike criteria, see below.

number of hidden layers and nodes should be able to approximate well non linearities of this kind, as noted before.

As is considered in survival analysis, in the estimation of two logit models and the ANN structure we include in the index function $\omega_i(x_{it-r};\theta)$ a concave function of duration time, d_{it}, given by $h(d_{it}) = \theta_1 \ln(d_{it}) + \theta_2 (\ln(d_{it}))^2$ to obtain predictions of P_{it} condition conditional on d_{it} (see also our Discussion in 2.1). This function is chosen among a set of alternative quadratic forms of $h(d_{it})$, by comparing the fit of the two logit models into the data using the BIC metric. The concave nature of $h(d_{it})$ reflects evidence in the literature that loan default does not often occur at the very early and/or very late stage of their maturity horizon (see, e.g., Deng et al. (2000) and Glennon, Nigro (2005))). This can be also justified by the graph of Figure 2B.

To choose the best structure (architecture) of the ANN method in terms of the number of layers and their associated neurons, we present a number of prediction performance and data fit metrics. These metrics can be also used to compare the optimal ANN structure chosen to the two parametric models considered. The data fit metrics include the minus log-likelihood function value (denoted as $-loglik$) obtained by the ML optimization procedure, the BIC and the Mcfaddens' coefficients of determination R^2 and R^2-adjusted.[11] Note that, for the evaluation of the relative performance of the alternative ANN structures examined, one advantage of the BIC metric is that it can select a more parsimonious structure, in terms of parameters (or layers-neurons), compared to AIC which penalizes less more complex structures. The prediction performance metrics include the *type I* and *II* misclassification errors of the default or non-default events, as well as values of the *AV(ROC)* and *integral(ROC)* metrics, based on the receiver operating characteristic (*ROC*) curve. These metrics are often used in practice to evaluate the overall prediction performance of binary choice models (see, e.g., Satchel, Xia (2008), for surveys). The *type I error* metric gives the error (in percentage terms) that a loan is predicted as non-defaulted, while it is actually defaulted. On the other hand, the *type II error* gives the error that a loan is predicted as defaulted, while it is actually non-defaulted. To calculate these errors, we transform the

[11]These coefficients are defined as follows: $R^2 = 1 - \frac{\log L(\theta)}{\log L(\theta_R)}$ and R^2-adjusted $= 1 - \frac{\log L(\theta)\text{-}K}{\log L(\theta_R)}$, where θ is the parameter vector of the likelihood and θ_R is that of a restricted version of the model that assumes zero values for all slope coefficients.

predicted probabilities of default events, P_{it}, obtained by the alternative methods into binary (point) forecasts (i.e., $Y_{it} = 1$, or $Y_{it} = 0$) based on a threshold value above (or below) which a loan is characterized as defaulted (or non defaulted), denoted as p^* (see **?**). More specifically, the *type I* and *type II errors* are defined as:

$$\textit{type I error}: \quad e_I\left(p^*\right) = \frac{\sum_i \sum_t Y_{it} \left(\Pr(Y_{it} = 1 | Y_{is} = 0) \leq p^*\right)}{\sum_i \sum_t Y_{it}}, \quad \text{for } s < t, \tag{11}$$

and

$$\textit{type II error}: \quad e_{II}\left(p^*\right) = \frac{\sum_i \sum_t (1 - Y_{it}) \left(\Pr(Y_{it} = 1 | Y_{is} = 0) > p^*\right)}{\sum_i \sum_t (1 - Y_{it})}, \text{ for } s < t,$$

respectively. We can choose the optimal value of p^* by minimizing the sum of errors: $e_I\left(p^*\right) + e_{II}\left(p^*\right)$ (see Hwang and Chu (2014)).

The *ROC* curve can show how well a model discriminates between events $Y_{it} = 1$ and $Y_{it} = 0$ at any threshold value $p^* \in [0, 1]$. The true proportion rate of the default event, at time t, is defined as

$$TPR_t = \frac{\sum_i Y_{i,t} \left[\Pr(Y_{it} = 1 | Y_{is} = 0) > p^*\right]}{\sum_i Y_{it}}, \text{ for } s < t \tag{12}$$

and the false proportion rate (denoted as FPR_t) is the *type II error*. At each t, the *ROC* curve is defined by the geometric locus of the points (FPR_t,TPR_t), for all threshold values p^*, while the *ROCAUC* measures the area under the *ROC* curve, giving us a single number of performance. For the model that discriminates better between the default and non default events, the value of the *ROCAUC* curve, at any t, should be higher and closer to unity which is its maximum value. To account for the dynamic feature of our dataset, we report the average value, over all time points t, estimates of the *ROCAUC* curve, denoted as *AV(ROC)*, as well as the sum of *ROCAUC* over all sample points t, denoted as *integral(ROC)*. For the best performing method, both the *AV(ROC)* and *integral(ROC)* metrics will reach their highest values of all the methods considered.

[INSERT TABLES 1A-1B]

Tables 1A-1B present values of the *AV(ROC)* and *BIC* metrics. These clearly show that we need more than one layer of neurons to fit well an ANN structure into the data. This is true independently of the activation function considered. For $r = 3$, both the *AV(ROC)* and

BIC indicate that, for the logit activation function, the best ANN structure is that with 3 hidden layers and 4, 2 and 1 neurons in each of these layers, respectively. For the $ReLu$, the best structure is that with 3 hidden layers and 3, 2 and 1 neurons in each of these layers, respectively. When we move to the case of $r = 12$, the optimal number of layers and neurons decreases for both the activation functions, implying that a less complex ANN structure is required to predict the PD, efficiently, over longer horizons. As it becomes apparent from the results of Table 1B, for $r = 12$, the best ANN structure is that of 2 layers with 2 and 1 neurons associated with each layer, respectively.

To further appraise the differences of the ANN method, in terms of prediction power and fitness into the data, across the lag structures $r = \{3, 12\}$ and the two activation functions, in Tables 2A-2B we present values of the *type I*, *type II errors*, *AV(ROC)* and *integral(ROC)* metrics, as well as the $-loglik$ value for the optimal ANN structures (layers-neurons) indicated by the results of Tables 1A-1B. The results of the tables for $r = 3$ demonstrate that, for all the above metrics, the differences between the two activation functions examined are quite narrow, with the logit function to be slightly superior in terms of the *type I* error. As r increases to 12, all the metrics imply an important deterioration in the fitness and prediction performance of the method, independently of the activation function. Note that, for $r = 12$, the *type I error* increases almost by 30%.

[INSERT TABLES 2A-2B]

Tables 3A-3B and 4A-4B present parameter estimates, and values of the fit and prediction performance metrics for the two logit models. This is done for the lag structures $r = \{3, 12\}$. Note that the tables present results for alternative specifications of the covariates of the models, with, or without, the loan-specific group of variables and with, or without, the qualitative dummies distinguishing the corporate structure effects on P_{it} (i.e., SP and $Corp$). The comparison of the fit and prediction metrics, reported in the tables, to those of the optimally chosen structure of the ANN method, given in Tables 2A-2B, can lead to interesting conclusions about the benefits in using the ANN method in predicting the PD.

First, as can be seen from Tables 2A, 3A and 4A, the ANN method clearly performs better than the two logit models, in terms of most of the metrics presented in the tables (namely, the *type I+type II error*, *type I error*, *AC(ROC)* and *integral(ROC)*, for the case of

$r = 3$. Note that the gains of the method are more profound for the *type I error* metric, which implies an improvement almost by 60% even relative to the skewed logit model which constitutes a flexible parameterization. Note that the improvement becomes almost 150% when the ANN method is compared to the logit model, often used in practice. These results highlight the usefulness of the ANN method in managing bank loan portfolios. The reduction of the *type I error* can avoid losses due to falsely predicting loan non-defaulting. Regarding the *type I+type II error* metric, the improvements of the ANN method relative to the skewed and standard logit models are of order 25% and 30%, respectively.

Second, for $r = 12$, the results of the tables (see Tables 2B, 3B, 4B) indicate that the prediction performance of the ANN method is more similar to that of the two logit models. This can be confirmed by most prediction and fit metrics reported in the tables. Note that, even for this case, the ANN method is the best performing method in terms of the *type I error* metric. However, in terms of the total, *type I +type II error*, the best model is the skewed logit.

Summing up, the above results show that there exist significant gains from employing the ANN method in predicting the PD relative to the two logit models considered. These gains are more apparent over short term horizons. This can be attributed to possible non-linearities and unobserved effects driven borrowers' decisions to default in short-run. For longer horizons, these effects seem to cease, implying that the decisions to default may be determined by simpler structures which are in line with the logit models.

[INSERT TABLES 3A-3B and 4A-4B

Regarding the relative performance of the two logit models, the results of the tables demonstrate the superiority of the skewed logit model. This can be justified by almost all the fit and prediction performance metrics reported in the tables. The estimates of parameter a, reported in Tables 4A-4B, show that P_{it} exhibits an asymmetric to the right response pattern of P_{it} to the covariates of the model. This means that P_{it} approaches to unity with a slower rate than the standard logit model. Note that this result is robust to the alternative covariates specifications and the different cases of r considered. Regarding the remaining parameters of the two logit models, the results of the tables indicate that the slope coefficients of covariates $x_{k,it}$ have the correct sign (see Section 5), for all k, and are significant at the 5% level. The positive sign of the effects of the qualitative dummy SP on P_{it} indicate that sole proprietorship has a positive effect on the PD, despite evidence in the literature for the opposite (see Wellalage and Locke (2015)). This may be attributed to the systemic risk conditions held in the economy, during the sample and the moral hazard incentives raised by the legal procedures protecting individual borrowers and corporations from foreclosures. Under such conditions, small business companies, even those managed by sole proprietors, may tend not to service their debt, by repeatedly exploiting allowance of debt payment moratorium and loan restructuring.

Finally, in Figures 3A and 3B we present estimates of the central and tail-tendency measures of the distribution of the PD conditional on the duration time of a loan d_{it}, denoted as denoted as $\hat{P}_{it}\left(d_{it}\right)$. These are based on the estimates of the PD by the ANN and logit methods, and it is done for $r = \{3, 12\}$, respectively. The central-tendency measures include the mean and median, while the tail-tendency include the lower (25%) and upper (75%) percentiles. These measures enables us to study the effects of d_{it} on P_{it} in depth, across the whole support interval of the distribution $\hat{P}_{it}\left(d_{it}\right)$. For the ANN structure, the estimates of $\hat{P}_{it}\left(d_{it}\right)$ are based on the in-sample optimal estimates of the ANN structure given by Table 2A for the logistic activation function. For the two logit models, they are based on the full specifications of them given by the 7th-8th columns of Tables 3A-3B and 4A-4B.

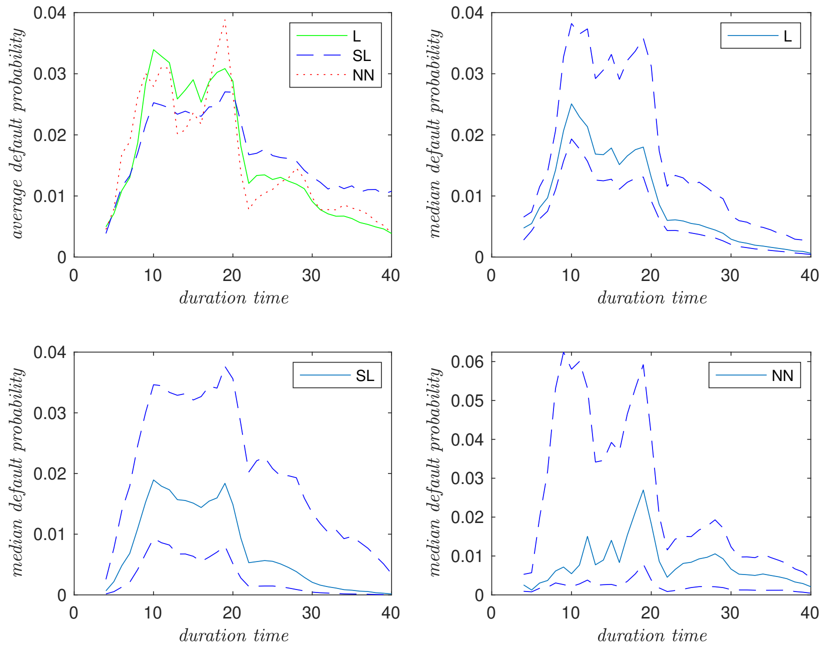

Figure 3A: Conditional on the duration PD functions ($s - t = 3$ months).

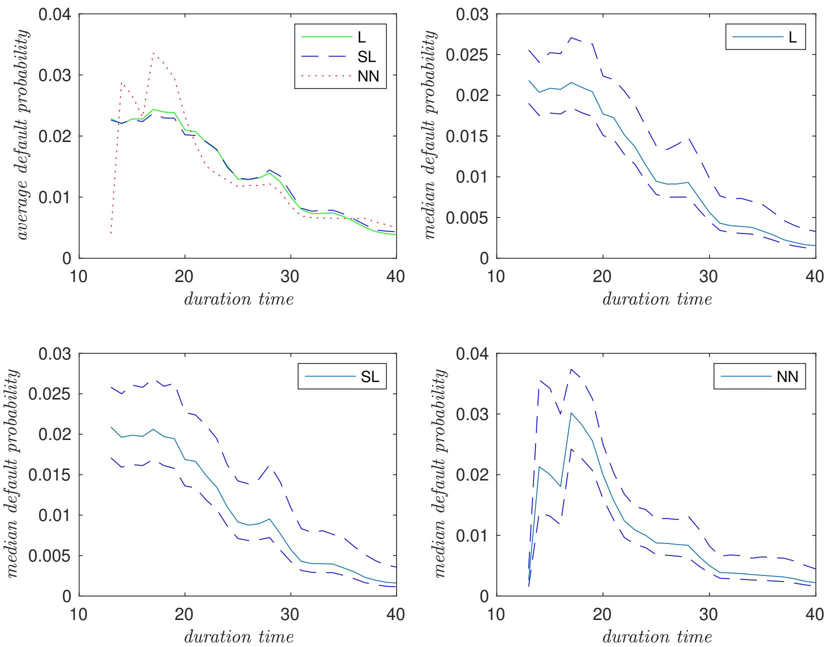

Figure 3B: Conditional on the duration PD functions ($s - t = 12$ months).

The inspection of Figures 3A and 3B implies that, for both cases of r, the distribution $\hat{P}_{it}(d_{it})$ obtained by the ANN method has an inverse-U shape. This is similar to that implied by the two logit models and it is consistent to shapes of $\hat{P}_{it}(d_{it})$ for other loan markets reported in the literature (e.g., for mortgages, see Dendramis et al (2018)). This result can be justified by all the graphs of the central-and tail-tendency measures of the distribution $\hat{P}_{it}(d_{it})$, reported in the figures. The pattern of all these measures tends to reach its highest values at relatively early duration intervals, which vary between 10 and 18 months, and then it exponentially declines. This pattern of $\hat{P}_{it}(d_{it})$ is also very close to that of the conditional default rates reported in Figure 2B, which is obtained nonparametrically from the data. The last result adds supports to the evidence that both the ANN method and the two logit models can sufficiently capture the true relationship between P_{it} and d_{it} implied by the data. Another interesting conclusion that can be drawn by the comparison of the graphs of Figures 3A and 3B is that $\hat{P}_{it}(d_{it})$ has a much more smooth pattern for $r = 12$ than $r = 3$. This may be attributed to the more complex pattern of the factors determining P_{it} over the shorter prediction horizons, such as $r = 3$, as also noted before.

6.2 The Generalized covariate effects on PD

To examine the relative importance of any covariates $x_{k,it-r}$, $k = 1, 2, ..., K$, on PD, in this section we present results for the generalized covariate effects $\Delta\hat{P}_{k,it}$. These are based on the bootstrap sampling method, described in Section 4, and the estimates of the three methods, presented in the previous section. For the ANN method, we use its specification with 3 hidden layers, and 4, 2 and 1 neurons in each of these layers, respectively, and the logit activation function, while for the logit models we use their full specifications whose estimates are reported in the 7th-8th columns of Tables 3A and 4A.[12] As for the lag structure r, we provide results for the case that $r = 3$ where all the methods exhibit better prediction performance and the ANN method outperforms the two logit models, according to our in-sample results. To better appraise the covariate effects for the most distressed loans, we focus on loan level data with duration $d_{it} = 12$ months. As Figure 2B shows, loans with age of around one year tend to have the highest default rates.[13]

[12]Note that, for the ANN method, similar results are obtained if the Re *Lu* is used as activation function.

[13]Note that most of these loans of our data set **refer to the sample period May-June 2015 where serious fears of a** Greek withdrawal from the Eurozone were raised, **the goverment** announced the bailout referendum on 27 June 2015, the banks closed and capital controls introduced.

Table 5 and Figure 4 present measures of the distribution of $\Delta\hat{P}_{k,it}$, for all $x_{k,it-r}$. In particular, the table presents the median, and the 25% and 75% percentiles of the distribution of $\Delta\hat{P}_{k,it}$, while Figure 4 presents the boxplots of $\Delta\hat{P}_{k,it}$. The results of the table and figure clearly show that, for all the methods, the effects of the ratios $deliq.amount/balance$, $no.delinq.buckets$ and $balance/installment$ on P_{it} are positive, while those of the ratio $payment/installment$ are negative. These results can be clearly supported by the distribution support interval of the boxplots of $\Delta\hat{P}_{k,it}$, and the median and percentile measures reported in the table. They are also consistent with the slope coefficients of the logit models, reported in Tables 3A-3B and 4A-4B, and the theoretical predictions made in Section 5.

A second conclusion that can be drawn from the results is that, in terms of the median the distribution of $\Delta\hat{P}_{k,it}$ which is a measure of central tendency, the covariate effects of the above ratios, reflecting behavioral attitudes of borrowers, are larger in magnitude than those of the macroeconomic covariates (namely, the inflation and real gdp growth rates), as well as those of the qualitative dummies *urban area* and SP. As can be seen from the boxplots of $\Delta\hat{P}_{k,it}$, the median of $\Delta\hat{P}_{k,it}$ for the macroeconomic variables is very close to zero and $\Delta\hat{P}_{k,it}$ has small dispersion around this value. Similar conclusions can be drawn for the covariates effects of *urban area* and SP on P_{it}. The above results are qualitative similar between the ANN method and the two logit models highlighting the robustness of the method to measure covariate effects. However, it is worthy to note that, in terms of magnitude, there are differences between the results. The ANN method provide larger values of $\Delta\hat{P}_{k,it}$ than the two logit models for the four ratios capturing behavioral attitudes of borrowers. This can be justified not only by the median, but also by the values of two quantiles reported in Table 5 for almost all of the ratios considered.

[INSERT TABLE 5]

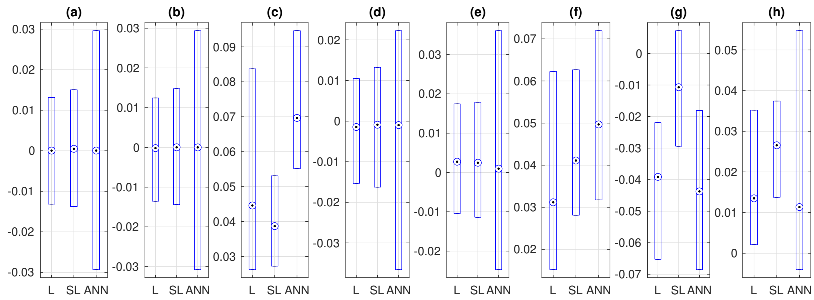

Figure 4: Covariate effects; (a) inflation rate (b) real gdp growth (c) *no.delinq.buckets* (d) Urban effects (e) *SP* (f) *delinq.amount/balance* (g) *payment/installment* (h) *balance/installment*. Also, L stands for logit, SL for skewed logit and ANN for Artificial Neural Network.

6.3 Out-of-sample results

In this section, we evaluate the out-of-sample forecasting performance of the ANN method and we compare the method to the logit and skewed logit models. For the ANN method, we adopt the optimal structure (no. of layers-neurons) implied by the in-sample results (see Tables 1A-1B, or Tables 2A-2B), and we examine the performance of both the logit and the *ReLu* activation functions. For the two logit models, again we rely on their full specifications reported in the 7th-8th columns of Tables 3A-3B and 4A-4B, including the whole set of covariates.

We obtain forecasts of the event that a loan i will default within a period (horizon) of $H = \{3, 12\}$ months ahead, defined as $Y_{i,t+h} = 1$, for $h \in \{1, .., H\}$. To obtain the forecasts for $H = 3$ (or 12), we assume that the lag structure is given as $r = 3$ (or 12), so that the forecast of the event $Y_{i,t+h}$ relies on the current t-time information of the sample. The calculation of the H-month ahead PD is based on the probability of survival of a loan i within the interval of H-months ahead, given as follows:

$$S_{it}(H) = \prod_{h=1}^{H} (1 - P_{i,t+h}), \text{ for } i = 1, .., N \tag{13}$$

where P_{it+h} is the PD given by (1). Given (13), we can obtain the PD of loan account i within the horizon of H-months ahead as $PD_{it}(H) = 1 - S_{it}(H)$. Based on $PD_{it}(H)$, we can classify if a loan i will default in the next H periods, when the condition $PD_{it}(H) \geq p^*$

holds, for $p^* \in [0,1]$. The optimal p^* can be found be minimizing the *type I+type II error*, as is described in Section 6.1.

Our forecasts rely on two recursive estimation approaches; the first uses an expanding window and the second a rolling over one. [14] The expanding window approach uses an initial data sample from 2014:03 until 2015:12 to obtain estimates of $P_{i,t+h}$ and provide forecasts of $PD_{it}(H)$ and loan default events. Then, we add to our initial sample window the next month of observations, obtain new estimates of $P_{i,t+h}$ and provide the subsequent set of $PD_{it}(H)$ and loan default events, within the next H months ahead. This procedure is sequentially repeated until the end of the sample (i.e., date 2017:06), leaving us an out-of-sample interval of one and a half years of observations for prediction performance evaluation. According to the rolling over window approach, we consider a rolling sample window of fixed size for the estimation. The first fixed window covers the period from 2014:03 to 2015:12. As in the expanding window approach, this window is used to obtain estimates of $P_{i,t+h}$ and $PD_{it}(H)$, and loan defaults. Next, it is moved one month forward and is used to obtain new estimates of $P_{i,t+h}$, and provide the subsequent values of $PD_{it}(H)$ and loan defaults. This procedure is repeated until the end of the sample. Note that, for both the above recursive estimation and forecasting approaches, the out-of-sample interval remains the same. This consists of 207396 observations, which constitute a relatively large sample size for efficient and reliable out-of-sample forecasting evaluation.

The results of the out-of-sample exercise are reported in Tables 6A-6B; Table 6A presents results for the ANN method and Table 6B for the two logit models. In each of these tables, we consider the cases of $H = \{3, 12\}$ and the two activation functions, for the ANN method. For evaluating the prediction performance of the methods on predicting default (or non default) events, the tables present the following metrics: the *type I* and *type II errors*, the *type I+type II error, and AV(ROC)* and *integral(ROC)* defined in Section 6.1. A general conclusion that can be drawn from the results of the tables is that the prediction performance of the ANN method is satisfactory and improves upon that of the two logit models, in terms of most of the metrics reported in the tables. This is true for both the expanding and rolling window forecasting approaches, and it holds for the two activation functions

[14] Note that, as it has been highlighted in the literature, the rolling window approach is considered as a more robust method to parameter instability (see e.g., **?** and Groen and Kapetanios (2016)).

considered.

The most improved metric by the ANN method is the *type I error*. This implies prediction gains that range over 100% in some cases. As we have previously discussed, controlling this error is important for bank loans management, seeking to minimize losses from defaults. Improvements are also observed by the ANN method for the metrics *AV(ROC)* and *integral(ROC)*, but these are less important than those of the *type I error*. The above conclusions mainly hold for the case of $H = 3$. For $H = 12$, the prediction gains of the ANN method relative to the two logit models reduce substantially. This is in accordance to our in-sample results. Moreover, the similar prediction performance of the ANN method between the in- and out-of-sample results, as well as between the expanding and rolling window forecasting approaches constitutes further evidence that the ANN structure (number of layers and neurons) chosen fits efficiently into the data, meaning that there are no overfitting problems due to over-parametirezation of the ANN structure (see West et al (1997)).

Finally, regarding the relative performance of the two logit models, the results of the tables are also consistent with the in-sample ones. They show that the skewed logit outperforms the logit, in terms of the type I error, $AV(ROC)$ and $integral(ROC)$ metrics, when $H = 3$. For $H = 12$, the differences in these metrics between the two models are eliminated.

[INSERT TABLE 6A, 6B]

7 Conclusions

Based on a large set of individual small business (SB) loan account data, this paper suggests a multilayer artificial neural network (ANN) - known as deep learning ANN - to predict the probability of loan default (PD) within the survival analysis framework. The ANN method has the advantage that it can consider complex structures in the data to capture the true factors driving loan default decisions. Furthermore, the interconnections between layers and neurons of the method can uncover hidden aspects capturing behavioral attitudes of borrowers or other institutional factors (e.g., legal procedures) influencing loan default decisions, which can lead to prediction gains. These factors are beyond the macroeconomic and credit risk fundamentals frequently used by parametric methods (e.g., the logit model) to predict the PD.

The paper provides a number of useful results highlighting the usefulness of employing the ANN method to predict loan defaulting. We show that this method constitutes a favorable alternative to the logit and skewed logit models. In particular, we find that the ANN method can lead to prediction gains in discriminating between default and non-default events, and to significantly reduce the type I misclassification error for default events. This can lead to significant loss reductions of bank loans portfolios. Our results are supported by in-sample estimates and out-of-sample forecasting exercises, over different horizons. They also demonstrate that most of the gains of the ANN method in predicting the PD are related to short term horizons. As the prediction horizon increases, the performance of the ANN method declines, but still compares favorably to the logit models for some metrics.

Finally, we show that loan-specific covariates, like the ratios of the delinquency amount of a loan over its total balance and loans payments over its installments reflecting behavioral attitudes of investors, constitute the main factors determining the PD. We provide clear cut evidence that the above covariates dominate the macroeconomic ones, such as the real gdp and inflation rate given at aggregate level, often used in the literature to predict the PD. These results are supported by a new measure of calculating generalized effects of a covariate on the PD over discrete changes of its values (e.g., different percentiles), while controlling for the effects of other covariates. This measure exploits the fact that the ANN method can provide accurate estimates of the PD and it can overcome identification issues of the ANN method encountered when calculating marginal effects. We derive the empirical distribution of the above covariate effects based on a bootstrap sampling method which takes into account possible interactions among the covariates determining the PD.

References

Altman Edward I., Sabato Gabriele. Modelling Credit Risk for SMEs: Evidence from the U.S. Market // Abacus. 2007. 43, 3. 332–357.

Barron A. R. Universal approximation bounds for superpositions of a sigmoidal function // IEEE Transactions on Information Theory. 1993. 39, 3. 930–945.

Behr Patrick, Foos Daniel, Norden Lars. Cyclicality of SME lending and government involvement in banks // Journal of Banking Finance. 2017. 77. 64 – 77.

Bergthaler Wolfgang, Kang Kenneth H, Liu Yan, Monaghan Dermot. Tackling Small and Medium Enterprise Problem Loans in Europe. III 2015.

Burr I.W. Cumulative frequency functions // The Annals of Mathematical Statistics. 1942. 13. 215–232.

Charalambakis E., Dendramis Y., Tzavalis E. On the determinants of NPLs: Lessons from Greece // Political Economy Perspectives on the Greek Crisis. 2017.

Cybenko G. Approximation by superpositions of a sigmoidal function // Mathematics of Control, Signals, and Systems (MCSS). XII 1989. 2, 4. 303–314.

Dendramis Y., Tzavalis E., Adraktas G. Credit risk modelling under recessionary and financially distressed conditions // Journal of Banking and Finance. 2009. 91. 160–175.

Dendramis Y., Tzavalis E., Varthalitis P., Athanasiou E. Predicting default risk under asymmetric binary link functions // International Journal of Forecasting. 2020. 36, 3. 1039 – 1056.

Deng Yongheng, Quigley John M., Order Robert. Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options // Econometrica. 2000. 68, 2. 275–307.

Dewatripont M., Tirole J. The Prudential Regulation of Banks. 1994.

Dietsch Michel, Petey Joël. The credit risk in SME loans portfolios: Modeling issues, pricing, and capital requirements // Journal of Banking Finance. 2002. 26, 2-3. 303–322.

Fausett Laurene. Fundamentals of Neural Networks: Architectures, Algorithms, and Applications. USA: Prentice-Hall, Inc., 1994.

Glennon Dennis, Nigro Peter. Measuring the Default Risk of Small Business Loans: A Survival Analysis Approach // Journal of Money, Credit, and Banking. 2005. 37. 923–947.

Greene W.H. Econometric analysis. 2012.

Gu Shihao, Kelly Bryan, Xiu Dacheng. Empirical Asset Pricing via Machine Learning // The Review of Financial Studies. 02 2020. 33, 5. 2223–2273.

Hornik Kurt, Stinchcombe Maxwell, White Halbert. Multilayer feedforward networks are universal approximators // Neural Networks. 1989. 2, 5. 359 – 366.

King Gary, Zeng Langche. Logistic Regression in Rare Events Data // Political Analysis. 2001. 9, 2. 137–163.

Koop Gary, Pesaran M.Hashem, Potter Simon M. Impulse response analysis in nonlinear multivariate models // Journal of Econometrics. 1996. 74, 1. 119 – 147.

Moro Andrea, Fink Matthias. Loan managers' trust and credit access for SMEs // Journal of Banking Finance. 2013. 37, 3. 927–936.

Rikkers Frieda, Thibeault Andre E. A Structural form Default Prediction Model for SMEs, Evidence from the Dutch Market // Multinational Finance Journal. 2009. 13, 3-4. 229–264.

Satchel Stephen, Xia Wei. 8 - Analytic models of the ROC Curve: Applications to credit rating model validation // The Analytics of Risk Model Validation. Burlington: Academic Press, 2008. 113 – 133. (Quantitative Finance).

Table 1A: Alternative layers-neurons specifications of the ANN model (lag structure $r = 3$ months)						
AV(ROC)	BIC	number of hidden layers	neurons in layer 1	neurons in layer 2	neurons in layer 3	neurons in layer 4
Activation Function: Logit						
0.839	38148.741	3	4	2	1	0
0.811	38549.508	2	5	1	0	0
0.828	38753.008	3	2	1	1	0
0.808	38769.953	2	4	1	0	0
0.825	38817.855	2	6	1	0	0
0.833	39112.385	3	3	1	1	0
0.816	39144.438	3	3	2	1	0
0.82	39598.628	2	3	1	0	0
0.815	39645.456	3	2	2	1	0
0.798	39856.441	2	2	1	0	0
0.814	40406.067	4	2	2	2	1
Activation Function: ReLu						
0.84	38480.532	3	3	2	1	0
0.825	39426.836	3	3	1	1	0
0.814	39508.526	2	3	1	0	0
0.815	39629.729	2	2	1	0	0
0.826	39659.436	2	4	1	0	0
0.825	39820.619	2	5	1	0	0
0.806	40494.894	3	4	2	1	0
0.8	40904.011	4	2	2	2	1
0.799	40907.275	3	2	2	1	0
0.799	41019.641	2	6	1	0	0
0.791	41419.352	3	2	1	1	0

Notes: The table presents values of the BIC and AV(ROC) metrics for alternative layers and neurons specifications of the survival ANN model, when the lag structure of the model is $r = 3$ months.

Table 1B: Alternative layer-neuron specifications of the ANN model (lag structure $r = 12$ months)						
AV(ROC)	BIC	number of hidden layers	neurons in layer 1	neurons in layer 2	neurons in layer 3	neurons in layer 4
Activation Function: Logit						
0.688	7042.504	2	2	1	0	0
0.577	7108.12	3	2	2	1	0
0.644	7173.595	2	3	1	0	0
0.679	7190.29	3	3	1	1	0
0.706	7220.828	2	6	1	0	0
0.686	7252.042	3	2	1	1	0
0.588	7262.05	3	4	2	1	0
0.688	7359.31	2	4	1	0	0
0.687	7370.413	3	3	2	1	0
0.683	7408.566	4	2	2	2	1
0.695	7454.724	2	5	1	0	0
Activation Function: ReLu						
0.683	7035.295	2	2	1	0	0
0.684	7120.492	3	3	2	1	0
0.685	7160.632	3	2	1	1	0
0.689	7269.147	4	2	2	2	1
0.685	7287.701	3	3	1	1	0
0.685	7288.955	2	3	1	0	0
0.677	7310.328	3	2	2	1	0
0.705	7354.47	2	6	1	0	0
0.686	7415.784	2	4	1	0	0
0.685	7457.006	3	4	2	1	0
0.698	7485.903	2	5	1	0	0

Notes: The table presents values of the BIC and AV(ROC) metrics for alternative layers and neurons specifications of the ANN model, when the lag structure of the model is $r = 12$ months.

Table 2A: Prediction and fit performance metrics for the ANN model (in-sample results- lag structure $r = 3$)		
	I: number of hidden layers=3, neurons $V_{l_j} = \{4, 2, 1\}$, $j = 1, 2, 3$, Act. function=logit)	II: number of hidden layers=3, neurons $V_{l_j} = \{3, 2, 1\}$, $j = 1, 2, 3$, Act. function= ReLu)
$-loglik$	-18718.147	-18965.287
BIC	38148.741	38480.532
no of parameters.	57	44
$p*$	0.011	0.013
type I + type II	0.362	0.385
type I error	0.093	0.111
type II error	0.269	0.274
AV(ROC)	0.839	0.84
integral(ROC)	28.491	28.508

Notes: The tables presents values of prediction and fit performance metrics for the best chosen specification of the ANN model with lag structure $r = 3$, based on the results of the BIC and AV(ROC) metrics reported in Table 1A.

Table 2B: Prediction and fit performance metrics for the ANN model (in-sample results- lag structure $r = 12$)		
I: number of hidden layers=2, neurons $V_{l_j} = \{2, 1\}$, $j = 1, 2$, Act. function=logit)		II: number of hidden layers=2, neurons $V_{l_j} = \{2, 1\}$, $j = 1, 2$, Act. function$\equiv$ ReLu)
$-loglik$	-3452.654	-3492.223
BIC	7042.504	7035.295
no of parameters.	73	25
$p*$	0.014	0.018
type I + type II	0.667	0.664
type I error	0.276	0.316
type II error	0.391	0.349
AV(ROC)	0.688	0.683
integral(ROC)	17.845	17.747

Notes: The tables presents values of prediction and fit performance metrics for the best chosen specification of the ANN model with lag structure $r = 12$ months, based on the results of the BIC and AV(ROC) metrics reported in Tables 1B.

Table 3A: Estimates of the logit model (lag structure $r = 3$)								
covariates\ coef. estimates	estimate	std error	estimate	std error	estimate	std error	estimate	std error
constant	-14.84**	(0.362)	-15.99**	(0.348)	-13.17**	(0.363)	-14.94**	(0.365)
urban area	0.039	(0.034)	0.038	(0.034)	-0.058	(0.034)	0.04	(0.034)
sole proprietorship (SP)	0.085**	(0.031)	0.084**	(0.031)	0.132**	(0.031)	0.224**	(0.069)
$\ln(d_{it})$	9.37**	(0.283)	10.28**	(0.269)	7.61**	(0.282)	9.37**	(0.282)
$\ln(d_{is}^2)$	-1.97**	(0.054)	-2.13**	(0.052)	-1.51**	(0.054)	-1.97**	(0.054)
no. delinq. buckets	0.079**	(0.002)	0.079**	(0.002)				
delinq. amount / balance	4.32**	(0.109)	4.29**	(0.109)				
payment / installment	-0.788**	(0.027)	-0.801**	(0.027)				
balance / installment	0.006**	(0.0001)	0.006**	(0.0001)				
no. delinq. buckets (SP)							0.081**	(0.002)
no. delinq. buckets ($Corp$)							0.077**	(0.002)
delinq. amount / balance(SP)							4.12**	(0.145)
delinq. amount / balance ($Corp$)							4.67**	(0.17)
payment / installment(SP)							-0.832**	(0.034)
payment / installment ($Corp$)							-0.713**	(0.044)
balance / installment (SP)							0.005**	(0.0001)
balance / installment ($Corp$)							0.007**	(0.0005)
inflation rate	0.009	(0.01)			0.008	(0.01)	0.009	(0.01)
real gdp growth rate	-0.249**	(0.021)			-0.234**	(0.021)	-0.249**	(0.021)
$-loglik$	-20732.13		-20797.88		-22942.52		-20727.87	
BIC	41601.76		41708.25		45972.54		41643.23	
no. parameters	11		9		7		15	
R^2	0.128		0.125		0.035		0.128	
R^2_{adj}	0.128		0.125		0.035		0.128	
$p*$	0.02		0.021		0.015		0.02	
type I + type II	0.489		0.484		0.729		0.49	
type I error	0.254		0.266		0.274		0.249	
type II error	0.235		0.218		0.455		0.241	
AV(ROC)	0.799		0.798		0.55		0.799	
integral(ROC)	27.12		27.09		18.68		27.11	

Notes: The table presents coefficient estimates and prediction and fit performance metrics for alternative specifications the logit model, when the lag structure is $r = 3$ months. We consider specifications excluding the group of macroeconomic variables and/or including dichotomous loan-specific covariates, categorizing the data across the corporate structure of the SBLs, i.e., sole proprietorship (denoted as SP) and corporations ($Corp$)). Standard errors are in parentheses. ** means significance at 5% level, * means significance at 1% level.

Table 3B: Estimates of the logit model (lag structure $r = 12$)

covariates\ coef. estimates	estimate	std error	estimate	std error	estimate	std error	estimate	std error
constant	-23.11**	(4.48)	-26.91**	(3.98)	-16.88**	(4.37)	-22.82**	(4.48)
urban area	0.103	(0.089)	0.104	(0.089)	0.022	(0.089)	0.107	(0.089)
sole proprietorship (SP)	0.236**	(0.081)	0.236**	(0.081)	0.331**	(0.081)	0.097	(0.16)
$\ln(d_{it})$	14.35**	(2.95)	16.76**	(2.63)	9.7**	(2.86)	14.12**	(2.95)
$\ln(d_{is}^2)$	-2.7**	(0.48)	-3.08**	(0.433)	-1.81**	(0.463)	-2.66**	(0.481)
no. delinq. buckets	0.074**	(0.006)	0.074**	(0.006)				
delinq. amount / balance	7.12**	(0.865)	7.07**	(0.862)				
payment / installment	-0.105	(0.057)	-0.114*	(0.058)				
balance / installment	0.003**	(0.001)	0.003**	(0.001)				
no. delinq. buckets (SP)							0.068**	(0.008)
no. delinq. buckets ($Corp$)							0.079**	(0.008)
delinq. amount / balance(SP)							7.9**	(0.945)
delinq. amount / balance ($Corp$)							8.08**	(2.5)
payment / installment(SP)							-0.055	(0.068)
payment / installment ($Corp$)							-0.112	(0.088)
balance / installment (SP)							0.005**	(0.001)
balance / installment ($Corp$)							0.002*	(0.001)
inflation rate	-0.038	(0.028)			-0.038	(0.028)	-0.039	(0.028)
real gdp growth rate	-0.097	(0.056)			-0.085	(0.055)	-0.098	(0.055)
$-loglik$	-3473.9		-3475.85		-3586.12		-3471.74	
BIC	7065.89		7048.33		7247.38		7104.51	
no. parameters	11		9		7		15	
R^2	0.056		0.055		0.025		0.056	
R^2_{adj}	0.053		0.053		0.023		0.052	
$p*$	0.018		0.019		0.018		0.021	
type I + type II	0.653		0.641		0.738		0.661	
type I error	0.302		0.335		0.323		0.431	
type II error	0.351		0.306		0.415		0.23	
AV(ROC)	0.688		0.688		0.536		0.687	
integral(ROC)	17.88		17.87		13.95		17.85	

Notes: See notes in Table 3A.

Table 4A: Estimates of the skewed logit model (lag structure $r = 3$)								
covariates\ coef. estimates	estimate	std error	estimate	std error	estimate	std error	estimate	std error
constant	-24.55**	(0.993)	-35.13	(45.13)	-11.8**	(0.312)	-16.76**	(0.333)
urban area	0.271**	(0.103)	0.414	(1.7)	-0.06	(0.039)	0.109	(0.093)
sole proprietorship (SP)	0.143*	(0.073)	0.191	(0.679)	0.136**	(0.032)	0.887**	(0.125)
$\ln(d_{it})$	18.81**	(0.753)	26.98	(31.86)	7.84**	(0.24)	12.6**	(0.852)
$\ln(d_{is}^2)$	-3.78**	(0.143)	-5.32	(6.23)	-1.55**	(0.045)	-2.6**	(0.331)
no. delinq. buckets	0.17**	(0.008)	0.227	(0.245)				
delinq. amount / balance	115.04**	(3.43)	176.77**	(43.39)				
payment / installment	-0.322**	(0)	-0.347	(1.22)				
balance / installment	0.037**	(0.001)	0.05	(0.051)				
no. delinq. buckets (SP)							0.121	(0.076)
no. delinq. buckets ($Corp$)							0.113	(0.153)
delinq. amount / balance(SP)							24.92**	(2.17)
delinq. amount / balance ($Corp$)							37**	(2.91)
payment / installment(SP)							-1.12**	(0.089)
payment / installment ($Corp$)							-0.247**	(0.076)
balance / installment (SP)							0.018**	(0.003)
balance / installment ($Corp$)							0.017**	(0.001)
inflation rate	-0.033	(0.027)			0.007	(0.031)	-0.012	(0.023)
real gdp growth rate	-0.493**	(0.062)			-0.243**	(0.042)	-0.328**	(0.092)
a (asymmetry)	0.008		0.005		0.2		0.031	
$-loglik$	-20429.24		-20476.88		-22944.18		-20399.89	
BIC	40995.98		41066.26		45975.85		40987.27	
no. parameters	11		9		7		15	
R^2	0.141		0.139		0.035		0.142	
R^2_{adj}	0.14		0.138		0.035		0.141	
$p*$	0.018		0.019		0.015		0.019	
type I + type II	0.449		0.446		0.729		0.471	
type I error	0.147		0.154		0.274		0.207	
type II error	0.303		0.292		0.455		0.264	
AV(ROC)	0.814		0.815		0.55		0.808	
integral(ROC)	27.65		27.66		18.68		27.44	

Notes: See notes in Table 3A.

Table 4B: Estimates of the skewed logit model (lag structure $r = 12$)								
covariates\ coef. estimates	estimate	std error	estimate	std error	estimate	std error	estimate	std error
constant	-19.96**	(5.31)	-24.76**	(3.18)	-15.35**	(1.2)	-19.77	(13.61)
urban area	0.136	(0.103)	0.135*	(0.068)	0.023	(0.09)	0.142	(0.111)
sole proprietorship (SP)	0.241*	(0.094)	0.242**	(0.077)	0.339**	(0.077)	0.021	(0.328)
$\ln(d_{it})$	14.02**	(3.48)	17.02**	(2.09)	9.61**	(0.768)	14.02	(8.8)
$\ln(d_{is}^2)$	-2.67**	(0.576)	-3.13**	(0.353)	-1.8**	(0.123)	-2.66	(1.46)
no. delinq. buckets	0.077**	(0.016)	0.077**	(0.011)				
delinq. amount / balance	16.92**	(2.6)	16.32**	(2.45)				
payment / installment	-0.042	(0.111)	-0.049	(0.063)				
balance / installment	0.006**	(0.001)	0.005*	(0.002)				
no. delinq. buckets (SP)							0.065**	(0.017)
no. delinq. buckets ($Corp$)							0.094**	(0.011)
delinq. amount / balance(SP)							19.77**	(3.42)
delinq. amount / balance ($Corp$)							14.72**	(4.24)
payment / installment(SP)							-0.032	(0.099)
payment / installment ($Corp$)							-0.078	(0.1)
balance / installment (SP)							0.008	(0.005)
balance / installment ($Corp$)							0.004	(0.008)
inflation rate	-0.048	(0.032)			-0.039	(0.023)	-0.049	(0.042)
real gdp growth rate	-0.11	(0.062)			-0.085	(0.055)	-0.11	(0.122)
a (asymmetry)	0.067		0.072		0.264		0.062	
$-loglik$	-3466.55		-3468.56		-3586.52		-3463.37	
BIC	7051.18		7033.75		7248.2		7087.77	
no. parameters	11		9		7		15	
R^2	0.058		0.057		0.025		0.058	
R^2_{adj}	0.055		0.055		0.023		0.054	
$p*$	0.019		0.021		0.018		0.018	
type I + type II	0.65		0.644		0.738		0.657	
type I error	0.352		0.41		0.323		0.356	
type II error	0.298		0.234		0.415		0.301	
AV(ROC)	0.686		0.686		0.536		0.688	
integral(ROC)	17.82		17.82		13.95		17.86	

Notes: See notes in Table 3A.

Table 5: Covariate effects on the PD									
	logit			skewed-logit			ANN		
	median	25%	75%	median	25%	75%	median	25%	75%
inflation rate	0	-0.0132	0.0131	0.0005	-0.0138	0.015	0.0001	-0.0293	0.0296
real gdp growth	-0.0002	-0.0135	0.0124	0	-0.0144	0.0148	0	-0.0308	0.0294
no. delinq. buckets	0.0446	0.0262	0.0837	0.0387	0.0272	0.0531	0.0697	0.0551	0.0946
urban area	-0.0014	-0.0153	0.0105	-0.0009	-0.0162	0.0132	-0.001	-0.0366	0.0223
SP	0.0027	-0.0105	0.0175	0.0025	-0.0114	0.0179	0.001	-0.0247	0.0361
delinq. amount / balance	0.0312	0.0152	0.0622	0.0411	0.0281	0.0627	0.0497	0.0317	0.0719
payment / installment	-0.0391	-0.0653	-0.0219	-0.0107	-0.0294	0.0072	-0.0437	-0.0685	-0.0181
balance / installment	0.0135	0.0021	0.0352	0.0266	0.0137	0.0374	0.0114	-0.004	0.0547

Notes: Estimates of the median, the 25% and 75% quantiles of the empirical distribution of the interdecile effects $\Delta\hat{P}_{k,it+r}$. This is done for the survival ANN model, the logit and skewed logit models, for the lag structure $r = 3$ months, on loan accounts with duration d_{it} =12 months.

Table 6A: Out-of-sample prediction performance of the ANN model								
	$H = 3$ months ahead				$H = 12$ months ahead			
	expanding window		rolling window		expanding window		rolling window	
Activation Function	logit	ReLu	logit	ReLu	logit	ReLu	logit	ReLu
type I+type II	0.472	0.452	0.455	0.541	0.547	0.602	0.542	0.556
type I	0.098	0.099	0.104	0.235	0.174	0.311	0.174	0.149
type II	0.374	0.353	0.352	0.306	0.373	0.292	0.368	0.406
AV(ROC)	0.815	0.818	0.82	0.764	0.714	0.699	0.709	0.705
integral(ROC)	12.196	12.238	12.286	11.466	2.889	2.839	2.86	2.863

Table 6B: Out-of-sample prediction performance of the logit and skewed logit models								
	$H = 3$ months ahead				$H = 12$ months ahead			
	expanding window		rolling window		expanding window		rolling window	
Model	logit	sk. logit	logit	sk. logit	logit	sk. logit	logit	sk. logit
type I+type II	0.553	0.533	0.554	0.515	0.576	0.576	0.575	0.574
type I	0.23	0.183	0.203	0.187	0.244	0.244	0.177	0.202
type II	0.323	0.35	0.351	0.327	0.332	0.332	0.398	0.373
AV(ROC)	0.778	0.785	0.783	0.8	0.694	0.694	0.698	0.699
Integral(ROC)	11.642	11.753	11.717	11.973	2.819	2.819	2.854	2.863