MODELAR LAS DIFICULTADES FINANCIERAS DE LAS NUEVAS MICROEMPRESAS USANDO MÁQUINAS DE SOPORTE VECTORIAL: UN ESTUDIO DE CASO

RESUMEN: A pesar del destacado papel que desempeña el microemprendimiento en el desarrollo económico y de la alta tasa de quiebra que tienen las nuevas microempresas en sus primeros años de vida, muy pocos estudios han diseñado un modelo para detectar las dificultades financieras de los microemprendedores. Además, debido a la ausencia de investigaciones, no se conoce nada acerca de si la información no financiera y las técnicas estadísticas no paramétricas mejoran la capacidad predictiva de estos modelos. Por tanto, este artículo proporciona un innovador modelo para detectar las dificultades financieras específicamente diseñado para las microempresas de nueva creación mediante el uso de máquinas de soporte vectorial (MSV) y empleando variables financieras, no financieras y macroeconómicas. Basados en una muestra de casi 5.500 de una Institución Microfinanciera (IMF) peruana, nuestros hallazgos muestran que la introducción de información no financiera relacionada con la zona en la que el emprendedor vive y localiza su negocio, la duración de la relación IMF-emprendedor, el número de préstamos concedidos por la IMF en el último año, el destino del préstamo y la opinión de los expertos sobre la probabilidad de que la nueva microempresa experimente problemas financieros, aumentan de manera significativa la precisión de nuestro modelo de detección de dificultades financieras. Además, los resultados revelan que los modelos construidos usando MVS superan los obtenidos por aquellos modelos que emplean el tradicional análisis de regresión logística.

PALABRAS CLAVE: modelo de dificultades financieras, microempresas de nueva creación, instituciones microfinancieras, América Latina, variables no financieras.

MODÉLISER LES DÉTRESSES FINANCIÈRES DES NOUVELLES MICRO-ENTREPRISES EN UTILISANT DES MACHINES DE SUPPORT VECTORIEL : UNE ÉTUDE DE CAS.

RÉSUMÉ : Malgré le rôle important que joue le micro-entreprenariat dans le développement économique, et le taux élevé d'échec des nouvelles micro-entreprises dans leurs premières années d'existence, très peu d'études ont élaboré un modèle pour détecter les difficultés financières des micro-entrepreneurs. De plus, étant donné l'absence de travaux de recherche nous ne savons aucunement si l'information non financière et les techniques non paramétriques améliorent la capacité prédictive de ces modèles. Par conséquent, cet article propose un modèle innovant pour détecter les détresses financières, spécialement conçu pour les micro-entreprises qui viennent d'être créées par l'utilisation de machines à vecteurs de support (MVS) et en utilisant des variables financières, non financières et macroéconomiques. Nous basant sur un échantillon de près de 5.500 micro-entrepreneurs d'une Institution Micro-Financière (IMF) péruvienne, nos résultats montrent que l'introduction d'informations non financières liées à la zone où l'entrepreneur vit et situe son affaire, à la durée de la relation IMF−entrepreneur, au nombre de prêts accordés par l'IMF au cours de la dernière année, à la destination du prêt et l'avis des experts sur la probabilité que la nouvelle micro-entreprise connaisse des problèmes financiers, augmentent de manière significative la précision de notre modèle de détection de difficultés financières. De plus, les résultats montrent que les modèles construits en utilisant des MVS dépassent ceux obtenus par les modèles qui utilisent l'analyse traditionnelle de régression logistique.

MOTS-CLÉS : Modèle de difficultés financières, micro-entreprises de création nouvelle, institutions micro-financières, Amérique Latine, variables non financières.

CASO PRÁTICO: A ANÁLISE DOS PROBLEMAS FINANCEIROS DA CRIAÇÃO DE MICROEMPRESAS COM A AJUDA DE MÁQUINAS DE VETORES DE SUPORTE

RESUMO: Apesar do destacado papel que o microempreendimento desempenha no desenvolvimento econômico e da alta taxa de falências que as novas microempresas têm nos seus primeiros anos de vida, poucos estudos têm projetado um modelo para detectar as dificuldades financeiras dos microempreendedores. Além disso, devido à ausência de pesquisas, não se sabe nada sobre se a informação não financeira e as técnicas estatísticas não paramétricas melhoram a capacidade preditiva destes modelos. Portanto, este artigo proporciona um inovador modelo para detectar as dificuldades financeiras especificamente projetado para as microempresas de criação recente mediante o uso de máquinas de vetores suportes (MVS) e utilizando variáveis financeiras, não financeiras e macroeconômicas. Baseados em uma amostra de quase 5.500 microempresas de uma micro-instituição financeira (IMF) peruana, encontramos que a introdução de informação não financeira relacionada com a região na qual o empreendedor mora e localiza o seu negócio, a duração da relação IMF- empreendedor, o número de empréstimos concedidos pela IMF no último ano, a destinação do empréstimo e a opinião dos peritos sobre a probabilidade de a nova microempresa ter problemas financeiros aumentam significativamente a precisão do nosso modelo de detecção de dificuldades financeiras. Além do mais, os resultados revelam que os modelos construídos utilizando MVS ultrapassam os obtidos por aqueles modelos que utilizam a tradicional análise de regressão logística.

PALAVRAS-CHAVE: Modelo de dificuldades financeiras, microempresas de criação recente, micro-instituições financeiras, América Latina, variáveis não financeiras.

Modeling the Financial Distress of Microenterprise Start-Ups Using Support Vector Machines: A Case Study

Antonio Blanco-Oliver
Department of Financial Economics and Operations Management (VPPI-US), Faculty of Economics and Business, University of Seville, Spain.
E-mail: aj_blanco@us.es

Rafael Pino-Mejías
Department of Statistics and Operational Research, Faculty of Mathematics, University of Seville, Spain.
E-mail: rafaelp@us.es

Juan Lara-Rubio
Department of Financial Economics and Accounting, Faculty of Economics and Business, University of Granada, Spain.
E-mail: juanlara@ugr.es

ABSTRACT: Despite the leading role that micro-entrepreneurship plays in economic development, and the high failure rate of microenterprise start-ups in their early years, very few studies have designed financial distress models to detect the financial problems of micro-entrepreneurs. Moreover, due to a lack of research, nothing is known about whether non-financial information and non-parametric statistical techniques improve the predictive capacity of these models. Therefore, this paper provides an innovative financial distress model specifically designed for microenterprise start-ups via support vector machines (SVMs) that employs financial, non-financial, and macroeconomic variables. Based on a sample of almost 5,500 micro-entrepreneurs from a Peruvian Microfinance Institution (MFI), our findings show that the introduction of non-financial information related to the zone in which the entrepreneurs live and situate their business, the duration of the MFI-entrepreneur relationship, the number of loans granted by the MFI in the last year, the loan destination, and the opinion of experts on the probability that microenterprise start-ups may experience financial problems, significantly increases the accuracy performance of our financial distress model. Furthermore, the results reveal that the models that use SVMs outperform those which employ traditional logistic regression (LR) analysis.

KEYWORDS: Financial distress model, microenterprise start-ups, microfinancial institutions, Latin American, non-financial variables.

Introduction

Since the failure rate of microenterprise start-ups in their primary years is extremely high—50% according to Headd (2003), and 60% using the figures from Phillips and Kirchhoff (1989)—many micro-entrepreneurs have traditionally experienced major difficulties in accessing funding, thus preventing them from carrying out their business projects. This funding constraint exists despite the leading role that entrepreneurship plays in economic development, especially in relation to the reduction of the unemployment rate thanks to the high percentage of start-ups launched by people who are unemployed. This role is therefore an essential factor for social and

economic cohesion. Moreover, the creation of microenterprises is considered by many researchers as one of the best ways to increase the innovation and competitiveness of an economy (Audretsch & Keilbach, 2004; van Stel, Carree, & Thurik, 2005; Wennekers & Thurik, 1999; Wennekers, van Stel, Thurik & Reynolds, 2005). The problem accessing funding is greater and more common for those micro-entrepreneurs with a lower socio-economic level, mainly due to their lack of assets with which to secure funding. For this reason, in recent decades, a new paradigm of support, known as the microfinance industry, has grown dramatically for those micro-entrepreneurs excluded from the traditional financial system in emerging countries[1]. Microfinance Institutions (MFIs) offer business advice and grant small loans (called micro-credits) to the poorest entrepreneurs, that is, to those entrepreneurs excluded from the traditional financial system. In order to undertake this important social and economic task, MFIs have to assume a high credit risk, since it is the poorest micro-entrepreneurs who apply for micro-credits, and approximately 50-60% of microenterprise start-ups fail to survive their first years of existence. For these reasons, MFIs should have access to a model for detecting potential financial problems in these new business projects. This model could be useful for both MFIs and micro-entrepreneurs—MFIs could determine the probability that a customer will have financial problems before making a micro-credit loan, while micro-entrepreneurs could detect, before embarking on a new business, whether it is likely to be successful or, in contrast, to have financial problems. Thus, the main objective of this paper is to create a financial distress model specially designed for those micro-entrepreneurs financially excluded from the traditional financial system. Furthermore, since recent literature on default prediction models (see, for example, Altman, Sabato & Wilson, 2010; Grunert, Norden & Weber, 2005) has highlighted the value added by non-financial variables to the accuracy performance in predicting bankruptcy, we also test whether the introduction of non-financial information in our model leads to a better prediction of the financial distress of micro-entrepreneurs. The effect of the macroeconomic context on the prediction of entrepreneurs' financial difficulties is also considered. In order to achieve these goals, we use a data set containing financial ratios, non-financial variables, and macroeconomic predictors on almost 5,500 applicants from a Peruvian MFI, and we develop three financial distress models using support

vector machines (SVMs). The first model uses only financial ratios as predictor variables, the second model also incorporates non-financial variables, while the third model considers all three kinds of variables: Financial, non-financial, and macroeconomic predictors. We find that the introduction of non-financial variables results in a major improvement in the prediction of micro-entrepreneurs' financial distress, and that the model which considers all three kinds of variables (financial, non-financial, and macroeconomic predictors) obtains the best result.

In this study, the current literature on the financial distress of start-ups is updated in two ways. First, a financial distress prediction model is specifically designed for financially excluded micro-entrepreneurs by using SVM. To the best of the authors' knowledge, there is no model in the existing literature that has been exclusively created for this kind of firm using these statistical techniques. Second, based on earlier work on the relevance of non-financial information in the borrower-bank relationship carried out by Berger and Udell (1998) and Stein (2002), we explore the value added by non-financial information, known in the literature as "soft information", in the accuracy performance of the proposed financial distress model. Previous research has found that the introduction of non-financial variables increases the performance of the models (for example, see Altman *et al.*, 2010; Grunert *et al.*, 2005). Nevertheless, while considerable research has proven the suitability of financial factors for predicting financial difficulties (for example, see Altman, 1968), the role of non-financial variables remains ambiguous due to the scarcity of empirical evidence on this issue.

The remainder of the paper proceeds as follows. Section 2 provides an overview of the extensive literature on financial distress models. In Section 3, the methodology employed in this study (SVM) is laid out in detail. Section 4 describes the sample employed and the predictors considered. In Section 5, several financial distress models for the smallest microenterprises are presented using SVM and logistic regression (LR). In Section 6, the models are applied and compared using the test data set. Finally, in Section 7, we discuss the conclusions offered by the results, and analyze future lines of research.

Financial Distress Prediction in Entrepreneurship: A Literature Review

Many studies have dealt with the prediction of the financial distress of firms, however only a few have focused on the case of micro-entrepreneurs, and no research has ever used a non-parametric methodology such as that

---

[1]   Although the microfinance sector emerged in the 1970s in Bangladesh, it soon began to develop in Latin America where it has acquired great economic and social importance. The microfinance industry currently has its greatest impact, both socially and economically, in Latin America, and has experienced its greatest development there.

employed here. The first default prediction model was developed by Beaver (1966). Using univariate analysis and a matched sample consisting of 158 firms (79 failed and 79 non-failed), he analyzed 14 financial ratios. Altman (1968) was the first author who applied the multiple discriminant analysis (MDA) in bankruptcy prediction. By selecting a matched sample of 66 manufacturing firms (33 failed and 33 non-failed) during the period 1946-1965, and examining 22 potentially predictive financial ratios classified into five categories (activity, solvency, leverage, liquidity, and profitability), he found that working capital/total assets, retained earnings/total assets, earnings before interest and taxes/total assets, market value of equity/book value of total liabilities, and sales/total assets are the most predictive ratios[2]. Ohlson (1980) developed the first corporate failure study by employing logistic regression on a sample of 2,000 non-failed and 105 failed industrial firms during the period 1790-1976. Focusing on small and medium-sized enterprises (SMEs), Altman and Sabato (2007) developed one of the most relevant models

specifically made for this size of firm. Their study compared the traditional Z-score model with two new models which considered other financial variables and used traditional logistic regression. On a panel of data of over 2,000 U.S. SMEs in the period 1994-2002, these authors found that the new models outperformed the traditional Z-score model by almost 30% in terms of predictive power. Focusing on the models made for financially excluded entrepreneurs (microfinance industry), the first model applied to microfinance was developed by Viganò (1993). He applied discriminant analysis to a sample that contained 100 cases and 53 predictor variables of a microfinance institution from Burkina Faso. Its main drawback was the small sample size—only 100 cases. Sharma and Zeller (1997) created a model for an MFI in Bangladesh, using a database with 868 borrowers. After applying a Tobit methodology based on maximum-likelihood estimation, 5 of the initial 18 variables were found to be significant. Reinke (1998) used a probit model to develop a credit score in a South African microfinance institution. This used a sample of 1,641 cases, and found 8 significant explanatory variables. Zeller (1998) designed a credit scoring model using the Tobit

---

2   This model is widely known as the Altman Z-score.

methodology. He used a sample of 168 MFI borrowers in Madagascar and found 7 significant variables out of the initial 18. Schreiner's (1999) model was implemented in Bancosol, an MFI from Bolivia. This used a sample of 39,956 borrowers and applied a binary logistic regression, incorporating the 9 original variables into the final model. Also in Bolivia, Vogelgesang (2003) presented two models for two different MFIs. He applied a random utility model to two datasets that contained 8,002 and 5,956 cases respectively. In Mali, Diallo (2006) also employed logistic regression to make a credit scoring model using a sample of 269 borrowers. He only obtained 5 significant variables in his model. Kleimeier and Dinh (2007) formulated a scoring model for Vietnam's retail banking sector, also using logistic regression. The sample contained 56,037 cases, and 17 of the 22 original variables were found to be significant. Finally, Rayo, Lara and Camino (2010) created a model for Peruvian FMIs employing logistic regression. Based on a sample of 16,157 cases, setting aside 25% of them to validate the model, they found 12 significant variables of the 41 initially considered.

In terms of the capacity of non-financial variables to increase the accuracy performance of the default prediction models, one of the most important research studies was carried out by Altman *et al.* (2010). Based on previous research undertaken by Altman and Sabato (2007), Altman *et al.* (2010) explored the effect of introducing non-financial information as predictor variables into the models developed by Altman and Sabato (2007). Employing a large sample (5.8 million) of sets of accounts of unlisted firms from the UK in the period 2000-2007, they found that non-financial information greatly increases—by approximately 13% in terms of the area under the receiver operating characteristics curve (AUC)—the default prediction power of risk models. Grunert *et al.* (2005) created several default prediction models using both financial and non-financial variables (age and type of business, sector, etc.). They concluded that the combination of financial and non-financial variables improved the accuracy performance of the developed models. Peel and Peel (1989) and Peel, Peel, and Pope (1986) also show, using a sample of SMEs from the UK, that the timing of submission of annual accounts is an indicator of financial failure. However, whereas the importance of financial factors is widely accepted because their impact is measurable, the relevance of non-financial factors is mainly considered in a holistic manner.

Only a few studies have focused on developing financial distress models specifically for microenterprise, and nothing is known about whether non-financial information improves the predictive capacity of these models due to the lack of research on this issue. Moreover, as shown above,

there are no research studies in the literature that develop distress models for entrepreneurs employing a non-parametric methodology such as SVM, as is undertaken in this study. We assert that a gap exists in the established body of knowledge in this field and therefore, in this paper, we present a financial distress model specifically designed for micro-entrepreneurs using SVM and employing both financial and non-financial variables.

Methodology

Financial distress problems lie within the scope of the more general and widely discussed classification problems. Historically, classification problems were addressed by employing parametric statistical techniques, mainly linear or quadratic discriminant analysis (LDA or QDA) and logistic regression (LR). However, the strict assumptions (linearity, normality, or equality of dispersion matrices) of these traditional statistical models, together with the pre-existing functional form relating response variables to predictor variables, limit their application in the real world. For these reasons, in recent years, non-parametric statistical techniques such as classification trees, neural network models, and support vector machines have been successfully applied to classification problems in several fields. Of these tools for pattern classification, one of the most powerful is the support vector machine (SVM), particularly in its non-linear transformation version. SVM is a family of supervised machine learning techniques, originally introduced by Vapnik (1998) and other co-authors (Boser, Guyon & Vapnik, 1992), with several extensions successively proposed. When used for a two-class classification problem where the set of binary-labeled training patterns is linearly separable, the SVM method separates the two classes with a hyper-plane that is maximally distant from the patterns ("the maximal margin hyper-plane"). If linear separation is not possible, the feature space is enlarged using basis expansions such as polynomials or splines. However, explicit specification of this transformation is not necessary, only a kernel function that computes inner products in the transformed space is required. We have fitted the SVM models with the svm function available in library e1071 of the R system (Dimitriadou, Hornik, Leisch, Meyer & Weingessel, 2006), which offers an interface to the award-winning C++ library, LIBSVM, by Chang and Lin (2011). The data set is described by n training vectors $\{x_i, y_i\}$, $i = 1, 2, \ldots, n$, where the p-dimensional vectors x_i contain the predictor features and the n labels $y_i \in \{-1, 1\}$ identify the class of each vector. From among the several variants of SVM existing in library e1071, in line with Meyer (2004), we have used C-classification with the Radial Basis Gaussian kernel function:

$$K(\mathbf{u},\mathbf{v}) = \exp\left(-\gamma\left\|\mathbf{u}-\mathbf{v}\right\|^2\right) \qquad (1)$$

The primal quadratic programming problem to be solved is:

$$\underset{\mathbf{w},b,\xi}{Min} \quad \frac{1}{2}\mathbf{w}^t\mathbf{w} + C\sum_{i=1}^{n}\xi_i \qquad (2)$$

$$y_i(\mathbf{w}^t\varphi(\mathbf{x}_i)+b) \geq 1-\xi_i$$

$$\xi_i \geq 0, \, i = 1,2,...,n$$

$C > 0$ is a parameter controlling the trade-off between margin and error, and $\sum_{i=i}^{n}\xi_i$ is an upper bound on the sum of distances of the wrongly classified cases to their correct plane. The dual problem is:

$$\underset{\alpha}{Max} \quad \frac{1}{2}\alpha^t Q\alpha - \mathbf{e}^t\alpha \qquad (3)$$

$$0 \leq \alpha_i \leq C, \quad i = 1,2,...,n$$

$$\mathbf{y}^t\alpha = 0$$

where e is the n-vector of all ones, and Q is a positive semidefinite matrix defined by $Q_{ij} = y_i y_j K(x_i, x_j)$, $i,j = 1,2,...,n$, where $K(x_i, x_j) = \phi(x_i)\phi(x_j)$ is the kernel function. No explicit construction of the non-linear mapping $\phi(x)$ is needed, which is known as the kernel trick. A vector x is classified by the decision function

$$sign\left(\sum_{i=1}^{n}y_i\alpha_i K(x_i, x) + b,\right) \qquad (4)$$

which depends on the margins $m_i = \sum_{i=1}^{n}y_i\alpha_i K(x_i, x) + b, \quad i = 1,2,...,n$. The greater the absolute value of the margin, the more reliable the computed classification. The exploratory analysis of the margins could help throw light on the SVM model (Furey *et al.*, 2000). The solution to the quadratic programming problem is global, and avoids the non-optimality of the training process of other models, as happens with neural networks.

Two parameters must be tuned: C and γ. We have adopted the suggestions of Meyer (2004) to select the parameters of the SVM model, and therefore a grid search for C was first defined over the set {1,10,20,30,40,50, 100,150,....,1000}. Second, a grid search for γ was conducted in the set {0.10,0.15,0.20,...,0.90}. This configuration of C and γ was performed by a 10-fold cross-validation mechanism with the function tune.svm in the e1071 library. One drawback of the SVM model is the need to correctly identify appropriate values for the required parameters as we have described. Another drawback is that although several schemes exist which attempt to interpret the final model, none of them is sufficiently clear.

With respect to the evaluation procedure models, the receiver operating characteristic (ROC) is used. Although the AUC is widely used to evaluate this kind of model, other classification accuracy measures employed in previous research, such as the Brier Score (BS)[3], correct classification rate, and Type I-Type II error rates, were also employed. It is well known that the cost associated with a Type I error (a micro-entrepreneur without financial problems misclassified as a borrower with financial problems) and a Type II error (a customer with financial problems misclassified as a micro-entrepreneur without financial problems) is significantly different. Generally, the misclassification costs associated with Type II errors are much higher than those associated with Type I errors, and we must therefore pay special attention to Type II errors.

Data and Variables

The Data Set

We used a data set of micro-credits from a Peruvian MFI (Edpyme Proempresa), containing customer information for the period 2003-2008 related to: (a) Financial ratios of each microenterprise business project; (b) non-financial information related to both personal characteristics (marital status, gender, etc.) and the characteristics of the financial operation (rate of interest, amount, etc.); and (c) variables related to the macroeconomic context within which the business project was undertaken. After eliminating missing and abnormal cases, 5,451 cases remained. From among these, 2,673 (49.03%) were cases with financial problems and 2,778 (50.97%) were without financial problems. In line with other studies (for example, see Schreiner, 1999), we used a definition stipulating that micro-entrepreneurs have financial problems when they delay payment for at least 15 days. In order to perform an appropriate test of the models, we randomly split our final data into two subsets: A training set of 80%, and a test set of 20%. The training sample contained a total of 4,361 cases (51.09% with financial problems and 48.91% without financial difficulties), and the test set was composed of 1,090 cases of which 50.46% had financial problems. The training set was employed to estimate the parameters of the models and the test set was used to test the accuracy performance of the model (for greater detail, see Hastie, Tibshirani & Friedman, 2009).

---

[3] Brier Score is the mean quadratic discrepancy between probability of distress and the target 0-1 variable (non distress/distress).

TABLE 1. Description of Financial, Non-financial, and Macroeconomic Variables

Variable	Description	Expected Estimator Sign (β)
Financial Ratios		
R1	Assets rotation: Income Sales / Total Assets	-
R2	Productivity: Gross Utility / Operating Costs	-
R3	Liquidity: Cash / Total Asset Liquidity	-
R4	Liquidity rotations: Cash / Income Sales x 360	+
R5	Leverage1: Total Liabilities / (Total Liabilities + Shareholders' Total Equity)	+
R6	Leverage2: Total Liabilities / Shareholders' Equity	+
R7	ROA: Net Income / Total Assets	-
R8	ROE: Net Income / Shareholders' Equity	-
Non-Financial Information		
Zone	Geographical location of the agency or branch. Dummy variable: (0) central zone, (1) outskirts	+
Old	Duration as a borrower of the MFI. Numeric variable	-
Previous_Loan_Grant	Previously granted credits. Numeric variable	-
Loan_Grant	Loans granted in the last year. Numeric variable	-
Loan_Denied	Previously denied loans. Numeric variable	+
Sector	Activity sector of the micro-business. Categorical variable: (0) commerce, (1) agriculture, (2) production, (3) service	+ / -
Purpose	Destination of micro-credit. Dummy variable: (0) work capital, (1) fixed asset	+
Mfi_Class	MFI customer classification Dummy variable: (0) normal customer, (1) customer with repayment problems of any sort	+
Total_Fees	Total number of fees paid in credit history. Numeric variable	-
Arrears	Number of arrears. Numeric variable	+
Ave_Arrear	Average (days) of customer default. Numeric variable	+
Max_Arrear	Number of days of major default. Numeric variable	+
Gender	Borrower gender. Dummy variable: (0) male, (1) female	-
Age	Age at time of application. Numeric variable	+ / -
Marital_St	Marital Status. Dummy variable: (0) single, (1) family unit	-
Employm_St	Employment Status of borrower. Dummy variable: (0) owner, (1) dependent	+ / -
Guarantee	Guarantee presented. Dummy variable: (0) sworn declaration, (1) real guarantee	+
Currency	Type of currency for loan granted. Dummy variable: (0) Nuevos Soles (PEN) (1) US Dollar ($)	+
Amount	Amount of micro-credit. Numeric variable	-
Duration	Number of monthly fees for applied loan. Numeric variable.	+
Interest_R	Monthly interest rate for micro-credit. Numeric variable	+
Forecast	Loan officer forecast: credit situation at expiration. Dummy variable: (0) without problems, (1) with problems	+
Macroeconomic Indicators		
GDP	Rate of increase of annual change of Gross Domestic Product (GDP) during loan term	-
CPI	Rate of increase of annual change of Consumer Price Index (CPI) during loan term	+
Empl_R	Rate of increase of annual change of variation of employment rate (ER) during loan term	-
ER	Rate of increase of annual change of variation of exchange rate (ER) PEN1-$ during loan term	+
IR	Rate of increase of annual change of interest rate (IR) during loan term	+
SEI	Rate of increase of annual change of stock exchange index (SEI) during loan term	-
Water	Rate of increase of annual change in cost of municipal water during loan term	+
Electricity	Rate of increase of annual change in cost of electricity during loan term	+
Phone	Rate of increase of annual change in cost of telephone consumption during loan term	+

Source: Own elaboration.

Description of Input Variables

Table 1, below, shows the input variables used in this study[4]. The input variables employed contain the various characteristics of borrowers, lenders, and loans. Numerous qualitative variables are considered in our study, since: (a) Schreiner (2000) suggests that the input variables of credit scoring forces the microfinance sector to be more qualitative and informal than traditional banks; and (b) recent literature has concluded that the inclusion of qualitative variables improves the prediction power of models. Moreover, since borrower default is closely related to the general economic situation, variables linked to the macro-economic context are also considered as input variables. With respect to the dependent variable, the financial distress of micro-entrepreneurs, this takes a value of 1 if the micro-entrepreneur has had financial problems, otherwise it takes a 0 value.

The first ratio indicates the number of times that income exceeds total assets. Therefore, we estimate that the ratio $R1$ is inversely related to the probability of suffering financial difficulties. Ratio $R2$ measures the relationship between the gross and operating costs of the microenterprise. As with the previous ratio, we expect that the sign of its coefficient is negative, since the higher the value of this ratio, the more solvent the income/loss of the firm, and the lower the probability of financial difficulties. The third financial ratio, $R3$, measures the liquidity of the microenterprise. Due to the design of this ratio, the higher its value, the lower the probability that a micro-entrepreneur will have financial difficulties. Therefore, the sign of the estimate is expected to be negative. The fourth financial ratio, $R4$, indicates the number of days the microenterprise takes to recover its treasury. In this case, the larger the value of this variable, the greater the likelihood of financial problems, and the expected sign of the estimator is therefore positive. The fifth financial ratio, $R5$, represents the percentage of liabilities that microenterprises have in their financial structure. We understand that a high level of liabilities inversely affects the ability of micro-entrepreneurs to pay. Consequently, a positive sign of the estimator of this variable is expected. The sixth financial ratio, $R6$, measures the ratio between the amount of debt and equity, and thus complements the information provided by the previous variable. We estimate that a high debt ratio results in an increase in the likelihood of financial problems, hence

the positive estimator. The seventh financial ratio, $R7$, measures the return on assets (ROA). A higher return on assets should help reduce the likelihood of financial problems and a negative sign of this variable estimator is therefore expected. The final financial variable considered, $R8$, measures the return on equity (ROE), that is, the return accrued by owners of the company. The greater the financial return of a microenterprise, the smaller its financial problems. We therefore consider that the sign of the estimator of this variable should be negative.

Customers who both live in a central area and situate their microenterprise in a central area usually run less risk of financial distress than those in rural areas. Therefore, the variable *Zone* is expected to have a positive sign in the estimator. The age of the MFI-customer relationship implies that the bank knows the payment history of a customer in detail, and this is why the variable *Old* is inversely related to the probability of suffering financial problems. The variables *Previous_Loan_Grant* and *Loan_Grant* are expected to have a negative sign in the estimator for the same reasons as for the variable *Old*; a lasting relationship with the financial institution involves the lender knowing all the risk inherent to the customer and also believing that this customer is reliable (Crook, Hamilton & Thomas, 1992). For customers with loans refused in the past, the risk of financial problems is greater. Thus, the sign of the variable *Loan_Denied* is considered to be positive. Since there is no previous reference that suggests a criterion for the consideration of a sector with more financial problems than others, the sign of the variable *Sector* remains undetermined. For the variable *Purpose*, we propose a positive sign, as we understand that the micro-credit provided for the acquisition of an asset implies a greater risk than a credit destined for working capital because the process of asset recovery through depreciation takes longer. Borrowers with any payment problems in the past (greater risk) take the value 1 in the variable *Mfi_Class*, thus, we consider that the sign of the estimator is positive. The higher the amount in fees the customer has paid, the greater the experience as a customer, and the lower the probability of suffering financial problems. Therefore, the sign of the estimator of the variable *Total_Fees* is negative. However, the variables *Arrears*, *Ave_ Arrears*, and *Max_ Arrears* are closely related to the probability of non-payment, and hence their estimators have positive signs. According to Schreiner (1999), women are more reliable at making payments than men. Consequently a negative sign is considered for the variable *Gender*. There is no empirical evidence about the relationship between the variable *Age* and the probability of financial problems, therefore the sign of the estimator for this variable cannot be determined. Customers who are

---

4   Table 1 also shows the expected sign of the relationship between each input variable and the probability of suffering financial problems. The statistical descriptions of all the input variables are shown in Appendix A and Appendix B. These statistics are presented for each group (financial problems and non-financial problems).

responsible for a family unit usually have better debt payment behavior than those who are single, that is, those without family obligations (Kleimeier & Dinh, 2007). For this reason, the variable *Marital_St* must have a negative estimator. A positive estimator is expected in the variable *Employ_St*, since customers who have some experience in the running of a microenterprise have a lower probability of financial distress than those who have only worked as an employee (that is, without any experience as micro-entrepreneurs). In microfinance, the reputation of the borrower is the main guarantee. Hence, Edpyme Proempresa asks for only a sworn statement of their property from those customers who rarely have problems fulfilling their payment obligations. On the other hand, real guarantees are demanded from both new customers and those who have had some problems with payments in the past. Therefore, the sign of the estimator of the variable *Guarantee* must be positive. A micro-credit granted in foreign currency (not in local currency) is affected by a risk in the rate of exchange and, for that reason, a positive sign is expected in the estimator of the variable *Currency*. Edpyme Proempresa only accepts a micro-credit request for high amounts if customers have paid their previous micro-credits without any problems. Therefore, micro-credits of high amounts correspond to old customers and good payers, since these customers have a lower probability of suffering financial problems than those with micro-credits whose amount is lower. Therefore, a negative sign is expected in the estimator of the variable *Amount*. The literature on credit risk widely supports the idea that borrowers who take out long-term loans run a greater risk of defaulting on their payments than those that take out short-term loans. Therefore, the coefficient of the variable *Duration* must be positive. The higher the interest rate of a financing source, the more difficulties the borrower has repaying it. Consequently, the variable *Interest_R* must have a positive estimator. Finally, we believe that an important variable, albeit totally subjective, is the risk analyst's opinion of the probability that a customer may have financial problems. This defined by the variable *Forecast*, and we expect a positive sign in its estimator.

We also introduced variables with information about the economic cycle since, according to Mester (1997), the absence of this kind of variable has historically implied a major limitation for financial distress models. Furthermore, as stated by Kim and Sohn (2010), the macroeconomic environment is a key factor that directly affects the payment behavior of any borrower. The macroeconomic variables under consideration were calculated through the following expression:

$$\Delta VM_{i,j} = \frac{VM_{i+j} - VM_i}{VM_i} \qquad (5)$$

where,

$\Delta VM_{i,j}$ is the variation rate of the considered macroeconomic variable,

VM is the considered macroeconomic variable,

i is the moment of the granting of the loan, and

j is the micro-credit term.

Financial Distress Models for Micro-Entrepreneurs

Since one of our goals is to analyze the value added by non-financial information and the macroeconomic context to the accuracy performance of the financial distress model for entrepreneurs, we developed a model in three steps, resulting three models. First, we only introduced like predictor variables in the financial ratios (Model 1). Next, we also considered the non-financial variables (Model 2). And finally, the macroeconomic predictors were also taken into account (Model 3). Additionally, it is very important to obtain a parsimonious default prediction model and, therefore, only statistically significant predictors should be considered. As SVM is a non-parametric statistical technique, it is not possible to determine the most significant independent variables or the sign of the relationship with respect to financial distress (for more details on the coefficient of the independent variables, see Table 1). Therefore, in line with prior literature (see for example, Chen, Härdle & Moro, 2011), before implementing the SVM, we created a parametric model (logistic regression) for selecting the significant variables and checking that the signs of the estimator of each significant variable followed our expectations. In this way we were also able to confirm the results obtained about the value added by non-financial and macro-economic information through the SVM model, and compare the accuracy performance of the parametric model (LR) with the non-parametric methodology (SVM) with regards to detecting financial distress in micro-entrepreneurs.

Therefore, three LR models were initially developed. Two main methods were used to select significant variables: Forward and backward stepwise selection procedures. When a forward stepwise selection procedure was implemented in model 1[5], *Income Sales/Total Assets* and *Total Liabilities/Shareholder's Equity* were the most significant

---

[5]   In this study, the cut-off rate chosen (0.50) for all the models is in terms of the level of test set default rate. However, the optimum cut-off value cannot be found without careful consideration of each peculiarity of each particular lender, such as tolerance of risk, profit-loss objectives, recovery process costs and efficiency, possible marketing strategies, etc.

financial ratios. Table 2 lists the coefficients and significance level of all the variables finally considered in each model. To explore whether non-financial information increases the accuracy performance of our model, the most relevant non-financial information was introduced into the previous model, resulting in a new model (Model 2). To this end, a forward stepwise selection procedure was also implemented, and through this, the *Old, Zone, Loan_Grant, Purpose,* and *Forecast* variables were selected. Finally, we created a third model incorporating the macroeconomic predictors (Model 3). In this case, only one macroeconomic variable, *ER*, was significant.

As shown in Table 2, all the slopes (signs) of the significant variables followed our expectations. The relevance of these variables in the failure of entrepreneurs can also be analyzed by looking at the absolute values of coefficients of each variable. *Total Liabilities/Shareholder's Equity* is the most relevant variable in the model which considers only financial ratios (Model 1), whereas *Forecast* is the most important variable in the models which introduce non-financial and macroeconomic predictors (Model 2 and Model 3, respectively).

Our findings are in line with the previous literature since no accepted financial theory of bankruptcy exists (Peat, 2007). In spite of abundant literature, there is an absence of a framework that clearly explains the relationships between companies' financial behavior, measured through financial ratios, and the failure of those companies. Across countries, a variety of accounting systems, economic conditions, funding structures, and tax codes may affect the predictive power of the same financial ratios. For these reasons, the literature identifies a large number of possible financial ratios as useful in predicting the failure of a firm. For example, Chen and Shimerda (1981) show that out of more than 100 financial ratios, almost 50% were found useful in at least one empirical study. However, after more than four decades of research, there is still no consensus on which financial ratios provide good predictors and why (Westgaard & Van der Wijst, 2001), although previous empirical research has found that a firm is more likely to go bankrupt if it is unprofitable, highly leveraged, and suffers liquidity difficulties (Myers, 1977). Therefore, economic logic and prior literature both support the relevance of the two significant financial ratios, as demonstrated above.

With respect to non-financial predictors, previous literature suggests that an adequate bankruptcy model made specifically for microenterprises should not be based solely upon financial ratios, due to the very limited availability of financial ratios for micro-entities (Berger & Udell, 1998).

Therefore, non-financial information must play a highly significant role for this type of firm. Prior literature has found that non-favorable audit reports (Peel & Peel, 1989), delays in submitting annual accounts (Ohlson, 1980; Whittred & Zimmer, 1984), or the existence of payment problems (Altman *et al.*, 2010; Wilson & Altanlar, 2013) are relevant variables for predicting the failure of a firm. However, empirical evidence in this field is not abundant. Our study provides five new non-financial variables, specifically developed for the microfinance industry, which were found to be useful predictors.

With respect to macroeconomic variables, the empirical evidence is scarce and, therefore, cannot provide convincing conclusions. Mensah (1984) demonstrated that different macroeconomic environments affect the accuracy performance of bankruptcy models. In contrast, Liou and Smith (2007) found that the macroeconomic information has a low ability to detect firms' financial distress. In our case, the macroeconomic variable exchange rate (*ER*) represents the value of the US dollar (USD) against the Nuevo Sol (PEN)—in other words, how many PEN are necessary to acquire a U.S. dollar. Thus, a decreased exchange rate shows an appreciation of the PEN against the dollar.

Figure 1 shows changes in the exchange rate over most of the time horizon that covers the credits of the analyzed portfolios. According to what happened at the end of the 1990s, a rise in the value of the exchange rates caused by an unfavorable international economic environment caused a strong national currency depreciation against the U.S. dollar. In contrast, from 2003, the exchange rate (PEN to USD) fell due to a favorable external macroeconomic environment. The positive international financial conditions favored abundant liquidity in the financial markets and low risk aversion from foreign investors. This all led to an excess of U.S. dollars in the Peruvian economy and,

FIGURE 1. PEN to USD Exchange Rate in Peru (1997–2007)
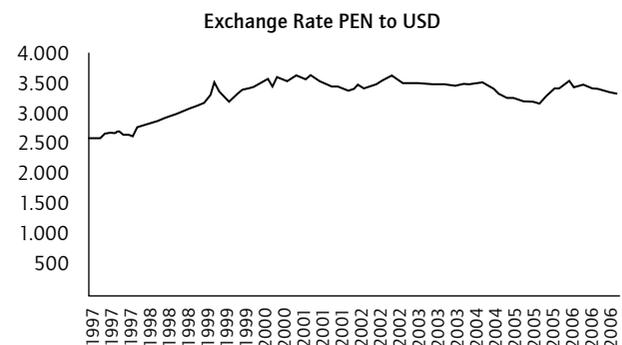
Source: Central Reserve Bank of Peru.

TABLE 2. Financial Distress Models for Entrepreneurs Using Logistic Regression

Ratio	Category	Model 1			Model 2			Model 3		
		Coefficient	Std. Error	Sig.	Coefficient	Std. Error	Sig.	Coefficient	Std. Error	Sig.
Income Sales/ Total Assets (R1)	Financial	-0.0802	0.0364	0.0279	-0.1507	0.0429	0.0004	-0.1263	0.0469	0.0071
Total Liabilities/Shareholder's Equity (R6)	Financial	0.4759	0.1020	0.0000	0.3961	0.1200	0.0009	0.5238	0.1340	0.0000
Old	Non-Financial				0.1322	0.0226	0.0000	0.1631	0.0251	0.0000
Zone	Non-Financial				0.5363	0.0688	0.0000	0.5126	0.0760	0.0000
Loan_Grant	Non-Financial				-0.1008	0.0154	0.0000	-0.1012	0.0166	0.0000
Purpose	Non-Financial				1.1310	0.1167	0.0000	1.0129	0.1292	0.0000
Forecast	Non-Financial				3.9393	0.2511	0.0000	4.4520	0.2561	0.0000
ER	Macroeconomic							0.2275	0.0097	0.0000
Intercept		0.0103	0.0477	0.8291	-0.5238	0.1039	0.0000	0.2111	0.1179	0.0734

Source: Own elaboration.

therefore, a weakening of the dollar against a strengthening of the PEN in a flexible exchange rate context.

Fluctuations in the exchange rate could also result in PEN losing value against the U.S. dollar; that is, the national currency could depreciate against the U.S. dollar. This, in turn, would increase the probability of default.

Once the variables that make up each of the three models were established, SVM models were implemented. We adopted the suggestions of Meyer (2004) in selecting the parameters of the SVM model, and therefore a grid search for C was first defined over the set $\{1,10,20,30,40,50, 100,150,....,1000\}$. Secondly, a grid search for γ was conducted in the set $\{0.10,0.15,0.20,...,0.90\}$. This configuration of C and γ was performed by a 10-fold cross-validation mechanism with the tune.svm function in the e1071 library. In the model that only considers financial ratios (Model 1), $C = 20$ and $\gamma = 0.3$. When both financial and non-financial variables are introduced (Model 2), $C = 10$ and $\gamma = 0.4$. Finally, in the model integrated for the three kind of variables (Model 3), $C = 10$ and $\gamma = 0.3$.

Testing and Comparison of Financial Distress Models

As explained above, part of our sample was retained for hold-out tests. This test sample contained 1,090 cases; of these, 540 (49.54%) had financial problems, and 550 (50.46%) had no financial distress. Table 3 summarizes the results in terms of AUC, Brier Score, correct classification rate, and Type I-Type II errors, for the three models tested

separately on the training and test samples (for more details about the models developed, see the tables of confusion in Appendix C and Appendix D).

As shown in Table 3, the model that includes both financial and non-financial information (Model 2) has a higher AUC and correct classification rate in the training and test sample than the model that considers only financial ratios (Model 1). Focusing on the results obtained for the out-of-sample set[6], the difference in the AUC between Model 2 and Model 1 is 21.61% and 24.44% for LR and SVM, respectively. This difference is 5.89% (LR) and 8.01% (SVM) when the Brier Score measure is considered. The results obtained by Type I-Type II errors and correct classification rate are found in the same way, and clearly show the superiority of the financial distress model that incorporates both financial and non-financial information (Model 2) over the model that exclusively introduces financial explicative variables (Model 1). This superiority is apparent in both statistical techniques used in this study (LR and SVM), which lends greater consistency and robustness to these findings. Therefore, in line with the findings in previous literature for SMEs (see, for example, Altman *et al.*, 2010; Grunert *et al.*, 2005), we suggest that the combined use of financial and non-financial variables increases the accuracy performance of the financial distress model designed exclusively for microenterprise start-ups. To test the effect

---

[6] We focus the discussion of our results on the out-of-sample set, since this is the best option to obtain unbiased results (Hastie *et al.*, 2009; Sobehart & Keenan, 2001).

TABLE 3. AUC, Brier Score, Correct Classification Rate and Type I-Type II Error Rates for LR and SVM Models

Models	Performance Measures	Training Sample			Test Sample		
		Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
LR Models	AUC	0.5460	0.7508	0.8405	0.5588	0.7749	0.8462
	Brier Score	0.2482	0.1953	0.1580	0.2478	0.1889	0.1515
	Correct classification rate	53.57%	69.07%	76.68%	53.12%	70.74%	78.72%
	Type I error rate	49.17%	17.20%	18.28%	48.33%	17.59%	18.33%
	Type II error rate	43.80%	44.07%	28.14%	45.45%	40.72%	24.18%
SVM Models	AUC	0.5667	0.8367	0.9099	0.5669	0.7913	0.8558
	Brier Score	0.2465	0.1747	0.1238	0.2427	0.1626	0.1163
	Correct classification rate	54.92%	76.20%	83.97%	60.09%	72.94%	81.47%
	Type I error rate	35.49%	15.52%	7.26%	35.37%	21.11%	13.15%
	Type II error rate	54.26%	31.73%	24.42%	44.36%	32.91%	23.82%

Source: Own elaboration.

of the macroeconomic environment on the prediction of the financial distress of microenterprise start-ups, the macroeconomic variable (*ER*) was also considered in Model 3. Our results suggest that the model which considers financial, non-financial and macroeconomic variables (Model 3) outperforms the model that only introduces financial and non-financial information (Model 2). In particular, the additional macroeconomic variable increases the AUC significantly (7.13% and 6.45% for LR and SVM, respectively), and reduces the Brier Score by 3.74% (LR) and 4.63% (SVM). These findings imply that the macroeconomic conditions also help detect the financial distress of micro-entrepreneurs. It is therefore suggested that introducing non-financial and macroeconomic information adds value to the financial distress model for microenterprise start-ups, with an improvement in terms of AUC of 21.61% (LR) and 24.44% (SVM) on introducing non-financial variables, and of 7.13% (LR) and 6.45% (SVM) on introducing the macroeconomic predictor. Moreover, as can be observed in Table 3, the introduction of non-financial and macroeconomic variables greatly decreases the occurrence of Type I-Type II errors by approximately 20%. This, in turn, has significant consequences for micro-entrepreneurs and MFIs. In the case of micro-entrepreneurs, a major Type I error curbs the creation of viable business projects; for MFIs, the greater the Type II error, the greater their losses.

Our findings also indicate that the non-parametric method implemented here (SVM), in terms of AUC, Brier Score, correct classification rate, and Type I-Type II error rates, outperforms the traditional logistic regression model (LR). This superiority is seen in the three models provided (Model 1, Model 2, and Model 3). Quantitatively, we found that the

SVM method improves the traditional LR technique in the range of 0.81-1.64% and 0.51-3.52% in terms of AUC and Brier Score, respectively. These empirical results confirm, for pattern classification problems, the theoretical superiority (principally, non-linear and non-parametric properties[7]) of the SVM technique over the parametric and widely used LR model, and this remains, moreover, consistent with previous literature (see for example, Chen *et al.*, 2011; Endorgan, 2013; Kim & Sohn, 2010).

In summary, we found, in line with the findings presented the literature for the case of SME bankruptcy models, that the introduction of non-financial and macroeconomic variables in our model not only increases its AUC and correct classification rate, but also decreases its Brier Score and Type I-Type II error rates. Therefore, we suggest that soft information contributes substantially towards increasing the accuracy performance of financial distress models built specifically for entrepreneurs. Moreover, due to the scarcity and often useless nature of financial information available to micro-entrepreneurs, our findings acquire more economic relevance than those found for general SMEs and large firms. Accordingly, lenders (MFIs) should collect and introduce non-financial information in their financial distress models for micro-entrepreneurs, since even a tiny improvement in predictive accuracy of the models can prove critical. In this sense, a 1% improvement in accuracy reduces losses in a large loan portfolio and can save

---

[7] Moreover, the SVM model obviates the need to assume strict assumptions of traditional statistical models and pre-existing functional forms that relate response variables to predictor variables that would otherwise limit application in the real world.

millions of dollars (West, 2000). In our case, the difference, in terms of the Type I and Type II error rates, between the model that considers the three kind of variables (financial, non-financial and macroeconomic, Model 3) and the model which only incorporates financial ratios (Model 1), is approximately 20%. Thus, the combination of financial ratios, and non-financial and macroeconomic predictive variables reduces lender losses drastically.

The financial distress models built here are advantageous for MFIs, micro-entrepreneurs, bank supervisors, and for economic development. First, lenders must control credit risk as effectively as possible, and they could therefore: (a) Reduce their capital requirements; (b) cut their costs and obtain a competitive advantage over those lenders that do not collect non-financial and macroeconomic variables; and (c) increase loan availability for micro-entrepreneurs, thereby promoting economic development. These models are also advantageous for micro-entrepreneurs, who could access more sources of finance thanks to an increase in the transparency of their financial information, since it reduces their problem of asymmetric information. Additionally, bank supervisors could benefit from the model since they would be supported in claiming to use a "mixed" credit scoring model which includes financial and non-financial variables to determine regulatory capital requirements (see Basel Committee on Banking Supervision, 2001). Finally, these models are advantageous for regional economic development since an increase in the creation of microenterprise start-ups constitutes an efficient way to reduce the unemployment rate, and increase the innovation, competitiveness, and growth of an economy (van Stel *et al.*, 2005; Wennekers & Thurik, 1999; Wennekers *et al.*, 2005).

Therefore, all stakeholders of microenterprises, particularly banks, creditors, and owners (micro-entrepreneurs), should carefully consider the results of this research for the detection of financial distress in this size of firm. In a restrictive environment, such as the one presented here, where the viable investment projects planned by micro-entrepreneurs often cannot be carried out by weak and cautious financial intermediaries, our financial distress model provides an innovative paradigm not only for mitigating the risk of failure in the microenterprise segment, but also for improving access to funding resources (mainly in the form of equity, bank debt, and commercial debt) for these types of firms. Additionally, the models developed here can be used by micro-entrepreneurs to analyze internal problems and control the performance of the company by anticipating insolvency situations and acting towards their solution.

Conclusions

Models for predicting the financial distress of micro-entrepreneurs are scarce in the literature. Furthermore, none of the models developed to date have assessed how the introduction of non-financial and macroeconomic variables affects their accuracy performance. Nor is there is any empirical evidence in previous literature on whether non-parametric techniques outperform classic parametric models in microenterprise start-ups. For these reasons, a financial distress model specifically designed for Peruvian micro-entrepreneurs was developed in this paper using support vector machines (SVM) and logistic regression (LR), and employing financial, non-financial, and macroeconomic information.

Our results suggest two relevant conclusions. First, this study confirms the findings of recent literature on SMEs (see for example, Altman *et al.*, 2010; Grunert *et al.*, 2005), which demonstrates the improvement in accuracy performance as a result of using non-financial information; thanks to the introduction of non-financial and macroeconomic information, accuracy performance was increased by more than 28% in terms of AUC. This improvement can be divided into the value added by the non-financial variables and macroeconomic predictor. In this sense, our results show an increase in the AUC of 21.61% (LR) and 24.44% (SVM) when non-financial information is introduced, and a further increment 7.13% (LR) and 6.45% (SVM) when the macroeconomic predictor is introduced. Therefore, the combined use of financial, non-financial, and macroeconomic variables for the development of financial distress models designed exclusively for microenterprise start-ups is recommended. Based on these results, Latin-American MFIs and micro-entrepreneurs should carefully consider the results of this research when granting a loan and when investing in a new business, respectively. These findings acquire even more economic relevance given the scarce and often useless financial information available on micro-entrepreneurs.

Second, our findings show that financial distress models can work well in the case of micro-entrepreneurs. In line with previous literature (see for example, Chen *et al.*, 2011; Endorgan, 2013; Kim & Sohn, 2010), we confirmed that SVMs obtain better results, in terms of AUC, Brier Score, correct classification rate, and Type I-Type II errors, than traditional logistic regression analysis (LR) in all the cases tested. In particular, our empirical results show that the non-parametric statistical method (SVM) improves the traditional LR technique in the range of 0.81-1.64%, and by

0.51-3.52% in terms of the AUC and Brier Score, respectively. Therefore, financial distress models designed for microenterprise start-ups, especially those developed under the SVM paradigm, constitute relevant tools that enable MFIs and micro-entrepreneurs to make better decisions about the viability of new business projects by reducing the high uncertainty associated with decision-making in this field, and, as a final consequence, allow an increase in the creation of viable business projects.

This study can help MFIs provide effective advice to micro-entrepreneurs about the viability of their business plans. As happens in the case of SMEs, this paper makes it clear that lenders should bear non-financial and macroeconomic information in mind before granting loans due to its relevance in the failure of microenterprises. Finally, our financial distress models should be applicable for microenterprises and MFIs in other Latin American countries since they operate in similar economic environments and socio-cultural contexts. It would also be possible to fit our failure models to other economic conditions by collecting the requisite information and data and replicating our analysis. The results of this study should be taken into account by all Latin American governments when creating economic policies to promote entrepreneurship, and by those entrepreneurs who are developing new businesses.

References

Altman, E. I. (1968). Financial ratios, discriminant analysis and prediction of corporate bankruptcy. *Journal of Finance*, *23*, 589-609.

Altman, E. I., & Sabato, G. (2007). Modeling credit risk for SMEs: Evidence from US market. *A Journal of Accounting, Finance and Business Studies (ABACUS)*, *43*(3), 332-357.

Altman, E. I., Sabato, G., & Wilson, N. (2010). The value of non-financial information in small and medium-sized enterprise risk management. *Journal of Credit Risk*, *6*(2), 95-127.

Audretsch, D. B., & Keilbach, M. (2004). Does entrepreneurship capital matter? *Entrepreneurship: Theory & Practice*, *28*(5), 419-429.

Basel Committee on Banking Supervision. (2001). *The new basel capital accord*. Consultative Document.

Beaver, W. (1966). Financial ratios as predictors of failure, empirical research in accounting: Selected studied. *Journal of Accounting Research*, *4*, 71-111.

Berger, A. N., & Udell, G. (1998). The Economics of small business finance: The roles of private equity and debt markets in the financial growth cycle. *Journal of Banking and Finance*, *22*(6-8), 613-673.

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). *A training algorithm for optimal margin classifiers*. In Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory. ACM Press, Pittsburgh, 144-152.

Chang, C. C., & Lin, C. J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*, 1-27.

Chen, K. H., & Shimerda, T. A. (1981). An empirical analysis of useful financial ratios', *Financial Management*, *10*(1), 51-60.

Chen, S., Härdle, W., & Moro, R. (2011). Modeling default risk with support vector machines. *Quantitative Finance,* *11*(1), 135-154.

Crook, J. N., Hamilton, R., & Thomas, L. C. (1992). *A comparison of discriminations under alternative definitions of credit default*. In Thomas, Crook & Edelman (Eds.), *Credit scoring and credit control* (pp. 217-245). Oxford: Oxford University Press.

Diallo, B. (2006). Un modele de 'credit scoring' pour une institution de microfinance Africaine: le cas de Nyesigiso au Mali. http://hal.archives-ouvertes.fr/docs/00/06/91/63/PDF/s16_05_06diallo.pdf

Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., & Weingessel, D. (2006). *e1071: Misc functions of the department of statistics (e1071), TU Wien*. R package version 1, 5-13. http://www.r-project.org

Endorgan, B. E. (2013). Prediction of bankruptcy using support vector machines: An application to bank bankruptcy. *Journal of Statistical Computation and Simulation*, *83*(8), 543-1555.

Furey, T. S., Cristianini, N., Duffy, N., Bednarski, D. W., Schummer, M., & Haussler, D. (2000). Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, *16*, 906-914.

Grunert, J., Norden, L., & Weber, M. (2005). The role of non-financial factors in internal credit ratings. *Journal of Banking and Finance*, *29*, 509-531.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*. New York: Springer Series in Statistics.

Headd, B. (2003). Redefining business success: Distinguishing between closure and failure. *Small Business Economics*, *21*, 51-61.

Kim, H. S., & Sohn, S. Y. (2010). Support vector machines for default prediction of SMEs based on technology credit. *European Journal of Operational Research*, *201*, 838-846.

Kleimeier, S., & Dinh, T. A. (2007). Credit scoring model for Vietnam's retail banking market. *International Review of Financial Analysis*, *16*(5), 471-495.

Liou, D., & Smith, M. (2007). Macroeconomic variables and financial distress. *Journal of Accounting, Business and Management*, *14*, 17-31.

Mensah, Y. M. (1984). An examination of the stationary of multivariate bankruptcy prediction models: A methodological study. *Journal of Accounting Research*, *22*(1), 380-395.

Mester, L. (1997). What's the point of credit scoring? *Business Review*. Federal Reserve Bank of Philadelphia, issue September, 3-16.

Meyer, D. (2004). *Support vector machines*. The interface to libsvm in package e1071. http://www.r-project.org

Myers, S. C. (1977). Determinants of Corporate Borrowing. *Journal Financial Economics*, *5*(2), 147-175.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, *18*(1), 109-131.

Peat, M. (2007). Factors affecting the probability of bankruptcy: A managerial decision based approach. *Abacus*, *43*(3), 303-324.

Peel, M. J., & Peel, D. A. (1989). A Multi-logit approach to predicting corporate failure- Some evidence for the UK corporate sector. *Omega International Journal of Management Science*, *16*(4), 309-318.

Peel, M. J., Peel, D. A., & Pope, P. F. (1986). Predicting corporate failure - Some results for the UK corporate sector. *Omega International Journal of Management Science*, *14*(1), 5-12.

Phillips, B., & Kirchhoff, B. (1989). Formation, growth and survival: Small-firm dynamics in the U.S. economy. *Small Business Economics*, *1*, 65-74.

Rayo, S., Lara, J., & Camino, D. (2010). A credit scoring model for institutions of microfinance under the Basel II normative. *Journal of Economics, Finance & Administrative Science*, *15*(28),89-124.

Reinke, J. (1998). How to lend like mad and make a profit: A microcredit paradigm versus the start-up fund in South Africa. *Journal of Development Studies*, *34*(3), 44-61.

Schreiner, M. (1999). The risk of exit for borrowers from a microlender in Bolivia. Center for Social Development, Washington University in St. Louis.gwbweb.wustl.edu/users/schreiner/.

Schreiner, M. (2000). Credit scoring for microfinance: Can it work? *Journal of Microfinance*, *2*, 105-118.

Sharma, M., & Zeller, M. (1997). Repayment performance in group-based credit programs in Bangladesh: An empirical analysis. *World Development*, *25*(10), 1731-1742.

Sobehart, J., & Keenan, S. (2001). Measuring default risk accurately. *Risk, 14*, 31-33.

Stein, J. C. (2002). Information production and capital allocation: Decentralized versus hierarchical firms. *Journal of Finance*, *57*, 1891-1921.

Van Stel, A., Carree, M., & Thurik, R. (2005). The effect of entrepreneurial activity on national economic growth. *Small Business Economics*, *24*(3), 311-321.

Vapnik, V. (1998). *Statistical Learning Theory*. New York: Wiley.

Viganò, L. A. (1993). Credit scoring model for development banks: An African case study. *Savings and Development*, *17*, 441-482.

Vogelgesang, U. (2003). Microfinance in times of crisis: The effects of competition, rising indebtedness, and economic crisis on repayment behavior. *World Development*, *31*(12), 2085-2114.

Westgaard, S., & Van der Wijst, N. (2001). Default probabilities in a corporate bank portfolio: A logistic model approach. *European Journal of Operational Research*, *135*(2), 338-349.

Wennekers, S., van Stel, A., Thurik, R., & Reynolds, P. (2005). Nascent entrepreneurship and the level of economic development. *Small Business Economics*, *24*(3), 293-309.

Wennekers S., & Thurik, R. (1999). Linking entrepreneurship and economic growth. *Small Business Economics*, *13*(1), 27-55.

West, D. (2000). Neural network credit scoring models. *Computers and Operations Research*, *27*, 1131-1152.

Whittred, G. P., & Zimmer, I. (1984). Timeliness of financial reporting and financial distress. *Accounting Review*, *59*(2), 297-295.

Wilson, N., & Altanlar, A. (2013). Company failure prediction with limited information: Newly incorporated companies. *Journal of the Operations Research Society*, *1-13*, doi:10.1057/jors.2013.31.

Zeller, M. (1998). Determinants of repayment performance in credit groups: The role of program design, intra-group risk pooling, and social cohesion. *Economic Development and Cultural Change*, *46*(3), 599-620.

Appendices

APPENDIX A. Statistical Description of Quantitative Independent Variables

Variable	Financial Problems		Non-financial Problems	
	Mean	Std. deviation	Mean	Std. deviation
R1	0.7637	0.8055	0.8436	0.8528
R2	3.9421	4.8284	3.8881	6.8548
R3	0.0683	0.0689	0.1448	3.2438
R4	0.1301	0.1368	0.1654	1.9812
R5	0.1421	0.1617	0.1196	0.1474
R6	0.2242	0.3227	0.1810	0.2789
R7	0.1531	0.1764	0.1771	0.2756
R8	0.1799	0.2015	0.2012	0.2911
Old	2.3468	1.5110	2.2397	1.5099
Previous_Loan_Grant	5.3900	5.0040	5.0600	4.6940
Loan_Grant	3.4600	2.3040	4.3400	2.3400
Loan_Denied	0.3200	0.5380	0.3300	0.5360
Mfi_Class	0.3500	0.4770	0.1100	0.3110
Total_Fees	36.1800	25.8510	31.7100	22.8390
Arrears	13.0400	10.7870	13.3400	11.1700
Ave_Arrear	8.0000	8.1510	6.8600	6.4340
Max_Arrear	20.2000	27.7650	16.5500	21.5030
Age	43.0175	10.6148	42.5628	10.4770
Amount	0.7338	0.6548	0.6458	0.5998
Duration	8.1100	4.7950	7.0300	3.5520
Interest_R	4.9242	0.9183	5.1255	0.8801
GDP	8.8985	29.7134	4.8139	26.3989
CPI	2.6377	2.2101	3.1247	2.1318
Empl_R	3.5702	10.6827	2.8671	9.6861
ER	-2.4123	4.4517	-5.5607	3.8899
IR	5.9631	13.9525	12.1717	11.7493
SEI	44.5991	32.4527	49.5322	33.3754
Water	2.4576	3.7483	3.1681	4.2243
Electricity	3.6054	12.2598	8.5162	10.4552
Phone	-7.1809	8.0019	-1.7179	3.8308

Source: Own elaboration.

APPENDIX B. Statistical Description of Qualitative Independent Variables

Variable	Categories	Financial Problems (%)	Non-financial Problems (%)
Zone Sector	Center	46.94	53.06
	Outskirts	55.84	44.16
	Commerce	48.53	51.47
	Agriculture	60.68	39.32
	Production	53.22	46.78
	Service	54.31	45.69
Purpose	Work Capital	47.07	52.93
	Fixed Asset	77.51	22.49
Gender Marital_St Employm_St	Male	51.32	48.68
	Female	50.71	49.29
	Single	50.73	49.27
	Family Unit	51.06	48.94
	Owner	50.81	49.19
	Dependent	70.73	29.27
Guarantee Currency Forecast	Sworn Declaration	58.50	41.50
	Real Guarantee	43.47	56.53
	PEN	89.30	92.10
	$	10.70	7.90
	Without problems	42.94	57.06
	With problems	97.27	2.73

Source: Own elaboration.

APPENDIX C. Table of Confusion Using Logistic Regression for Training and Test Samples

Training Sample			
Model 1		Classified class	
		0 (good credit)	1 (bad credit)
Actual class	0 (good credit)	1,084 (50.83%)	1,049 (49.17%)
	1 (bad credit)	976 (43.80%)	1,252 (56.20%)
Correct classification rate: 53.57%			
Model 2		Classified class	
		0 (good credit)	1 (bad credit)
Actual class	0 (good credit)	1,766 (82.80%)	367 (17.20%)
	1 (bad credit)	982 (44.07%)	1,246 (55.93%)
Correct-classification rate: 69.07%			
Model 3		Classified class	
		0 (good credit)	1 (bad credit)
Actual class	0 (good credit)	1,743 (81.72%)	390 (18.28%)
	1 (bad credit)	627 (28.14%)	1,601 (71.86%)
Correct classification rate: 76.68%			

Test Sample			
Model 1		Classified class	
		0 (good credit)	1 (bad credit)
Actual class	0 (good credit)	279 (51.67%)	261 (48.33%)
	1 (bad credit)	250 (45.45%)	300 (54.55%)
Correct classification rate: 53.12%			
Model 2		Classified class	
		0 (good credit)	1 (bad credit)
Actual class	0 (good credit)	445 (82.41%)	95 (17.59%)
	1 (bad credit)	224 (40.72%)	326 (59.28%)
Correct classification rate: 70.74%			
Model 3		Classified class	
		0 (good credit)	1 (bad credit)
Actual class	0 (good credit)	441 (81.67%)	99 (18.33%)
	1 (bad credit)	133 (24.18%)	417 (75.82%)
Correct classification rate: 78.72%			

Source: Own elaboration.

APPENDIX D. Table of Confusion Using Support Vector Machines for Training and Test Samples

Training Sample			
Model 1		Classified class	
		0 (good credit)	1 (bad credit)
Actual class	0 (good credit)	1,376 (64.51%)	757 (35.49%)
	1 (bad credit)	1,209 (54.26%)	1,019 (45.74%)
Correct classification rate: 54.92%			
Model 2		Classified class	
		0 (good credit)	1 (bad credit)
Actual class	0 (good credit)	1,802 (84.49%)	331 (15.52%)
	1 (bad credit)	707 (31.73%)	1,521 (68.27%)
Correct classification rate: 76.20%			
Model 3		Classified class	
		0 (good credit)	1 (bad credit)
Actual class	0 (good credit)	1,978 (92.74%)	155 (7.26%)
	1 (bad credit)	544 (24.42%)	1,684 (75.58%)
Correct classification rate: 83.97%			

Test Sample			
Model 1		Classified class	
		0 (good credit)	1 (bad credit)
Actual class	0 (good credit)	349 (64.63%)	191 (35.37%)
	1 (bad credit)	244 (44.36%)	306 (55.64%)
Correct classification rate: 60.09%			
Model 2		Classified class	
		0 (good credit)	1 (bad credit)
Actual class	0 (good credit)	426 (78.89%)	114 (21.11%)
	1 (bad credit)	181 (32.91%)	369 (67.09%)
Correct classification rate: 72.94%			
Model 3		Classified class	
		0 (good credit)	1 (bad credit)
Actual class	0 (good credit)	469 (86.85%)	71 (13.15%)
	1 (bad credit)	131 (23.82%)	419 (76.18%)
Correct classification rate: 81.47%			

Source: Own elaboration.