PREDICTING NON-LIFE INSURER’S INSOLVENCY USING NON-KERNEL FUZZY QUADRATIC SURFACE SUPPORT VECTOR MACHINES

YE TIAN
School of Business Administration and Collaborative Innovation Center of Financial Security
Southwestern University of Finance and Economics
Chengdu 611130, China

WEI YANG
School of Insurance and Collaborative Innovation Center of Financial Security
Southwestern University of Finance and Economics
Chengdu 611130, China

GENE LAI
Department of Finance
University of North Carolina at Charlotte
Charlotte, NC, 28223, USA

MENGHAN ZHAO
School of Business Administration
Southwestern University of Finance and Economics
Chengdu 611130, China

(Communicated by Shu-Cherng Fang)

Abstract. Due to the serious consequence caused by insurers’ insolvency, how to accurately predict insolvency becomes a very important issue in this area. Many methods have been developed to do this task by using some firm-level financial information. In this paper, we propose a new approach which incorporates several macroeconomic factors in the model and applies feature selection to eliminate the bad effect of some unrelated variables. In this way, we can obtain a more comprehensive and accurate model. More importantly, our method is based on the state-of-the-art non-kernel fuzzy quadratic surface support vector machine (FQSSVM) model which not only performs superiorly in prediction, but also becomes very applicable to the users. Finally, we conduct some numerical experiments based on the real data of non-life insurers from USA to show the predictive power and efficiency of our proposed method compared with other benchmark methods. Specifically, in a reasonable computational time, FQSSVM has the most accurate prediction rate and least Type I and Type II errors.

2010 Mathematics Subject Classification. 90B90, 90B50.

Key words and phrases. Non-life insurer, predicting insolvency, macroeconomic factors, feature selection, non-kernel FQSSVM.

Tian’s research has been supported by the Chinese National Science Foundation #11401485 and # 71331004. Yang’s research has been supported by the Sichuan soft science research project #2017ZR0294.
1. **Introduction.** It’s well known that insurers’ insolvency has a severely adverse influence on owners, banks, auditors, policyholders, stakeholders and even the general public. Note that more than 800 insurance companies in USA have been declared insolvent since 1970s. Thus, to protect the general public against the consequences of insurers’ insolvency and minimize the responsibilities for management and auditors, insurance commissioners need to identify which insurer has financial distress or trouble and subsequently take timely regulatory intervention. In this way, the insolvency may be prevented in time, the resulting cost can be maximally reduced and its effects on the whole economic system can be prevented or lessened. Therefore, how to accurately predict insurers’ insolvency becomes an important issue and has drawn increasing interest over these years.

Three decades ago, the National Association of Insurance Commissioners (NAIC) built the Insurance Regulatory Information System (IRIS) to provide early warnings of insolvency based on financial indicators. Two prominent solvency surveillance mechanisms/tools, risk-based capital (RBC) requirements and financial analysis solvency tools (FAST), were developed as a response to a rash of insolvencies during the 1980s. Note that RBC is derived from a formula which uses weighted risk-related items in the insurer’s financial conditions and has been used as a financial solvency indicator in insolvency studies [12]. The FAST system is more dynamic and employs more complicated ratio analysis than RBC requirements [8]. Besides, Several rating agencies have arisen and built their own systems, such as A.M.Best’s Ratings which combines qualitative and quantitative factors [3].

Prior researches have contributed a lot on insolvency prediction. A main stream of studies have extensively focused on firm-level variables and developed various insolvency prediction models. These studies only used cross-sectional and firm-specific data during relatively short sample periods [17, 22]. Besides, many researches have demonstrated that incorporating the explanatory variables related to changes in macroeconomic condition and insurance industry condition can significantly improve the qualities of prediction models [2]. Compared with previous models, these new models can avoid the volatility introduced by temporal trends during a long period. Therefore, they are more generalized and accurate in prediction while with fewer false positives [33].

Most approaches to insolvency study in insurance companies are statistical ways which use financial ratios as explicative variables, such as multivariate discriminant analysis (MDA) [14], logistic regression (LR) [5], stepwise logistic regression (SLR) [33], recursive partitioning model (RPM) [4] and nonparametric methods (NPM) [14]. However, this kind of variables does not usually satisfy statistical assumptions and hence the obtained results could be erroneous [19]. In addition to the statistical methods, some researchers also developed some operational and pattern recognition methods to predict insolvency, such as neural networks [7, 15], rough set [10] and genetic programming [21]. Though these methods achieve better results (less error in the classification) than traditional statistical approaches, their black-box character still make them difficult to interpret and the corresponding results cannot be clearly analyzed and related to the variables for discussion [23].

It is worth pointing out that support vector machine (SVM), a relatively new machine learning technique, has attracted a great deal of attention because of its promising features and excellent generalization performance on a wide range of problems [20]. SVM is an optimization-based binary classification technique which is easy to be analyzed mathematically. It aims to find a hyperplane that separates
the training points into two classes, with a maximum level of separation. Compared with other methods, SVM is more theory-driven, data-driven, distribution-free and robust [29]. In the last decade, SVM approach has been widely applied in insolvency prediction to improve the performance [19, 23, 26, 30, 31, 32]. In general, the numerical results demonstrate that SVM outperforms neural networks and statistical methods in predicting financial failures [9]. However, a kernel function is always necessary for the nonlinearly separable data sets in the classical SVM model [25]. It is used to map each training point to a higher dimensional space, where a hyperplane is then found to divide all data into two classes. Note that the performance of the SVM model largely depends on the choice of the kernel function and its corresponding parameters [6]. However, there is no universal rule to determine a suitable kernel function for any data set. Therefore, finding a good kernel function and its corresponding proper parameters is a hard work for the users and damages the applicability of the method.

To overcome this defect, we applied a state-of-the-art technique, non-kernel fuzzy quadratic surface support vector machine (FQSSVM), to predict insolvency in non-life insurance companies. This model was developed recently and the corresponding experimental results strongly demonstrated its great effectiveness for providing more accurate classifications than the classical SVM models [18, 28]. Since irrelevant input features may mess with the underlying system structure and then reduce the classification accuracy, we also investigate the feature selection in our prediction model. Note that feature selection aims to find a subset of input features which performs as well as all available features. Several researchers have noticed this problem and proposed some algorithms for searching optimal subsets of predictors [9, 11]. Finally, we conduct some numerical experiments based on the real data of non-life insurers from USA to show the predictive power and efficiency of our proposed method compared with other benchmark methods. Specifically, in a reasonable computational time, FQSSVM has the most accurate prediction rate and least Type I and Type II errors.

The contributions of this research include the following several aspects. First, FQSSVM applies a more general and effective classification method which avoids searching a proper kernel function and corresponding parameters in classical SVM models. Besides, fuzzy weight is introduced in the model to handle the imbalanced training points. Second, it is the first work to incorporate firm-level and macroeconomic information together to provide a more comprehensive and accurate model for predicting insolvency in non-life insurance companies. Third, it considers feature selection to exclude those ineffective variables in the model. Fourth, FQSSVM achieves the best prediction rate with least Type I and Type II errors. Finally, it demonstrates a big potential to be a useful tool for regulators and insurers in real applications and sheds some light on future researches.

The rest of the paper is organized as follows. We briefly review the classical SVM model and introduce the new FQSSVM model in Section 2. Then we explain the test data used in the numerical experiment and do the feature selection in Section 3. Section 4 shows some benchmark prediction methods in comparison and Section 5 shows the numerical results. At last, we summarize the paper in Section 6.

Now we introduce the notations used in this paper. The set of real numbers, the \( n \)-dimensional vector space, the \( n \)-dimensional nonnegative vector space and the space of \( n \times n \)-dimensional matrices are denoted as \( \mathbb{R} \), \( \mathbb{R}^n \), \( \mathbb{R}^n_+ \) and \( \mathbb{R}^{n \times n} \), respectively. Then for a vector \( \mathbf{x} \in \mathbb{R}^n \), its \( i \)-th component is \( x_i \), the 2-norm of \( \mathbf{x} \).
is denoted as \(\|x\|\). Moreover, \(\text{Diag}(x)\) is an \(n \times n\) matrix where its \(i_{th}\) diagonal element is \(x_i\). For a matrix \(M \in \mathbb{R}^{n \times n}\), the element in the \(i_{th}\) row and \(j_{th}\) column is denoted as \(M_{ij}\). For two matrices \(A, B \in \mathbb{R}^{n \times n}\), \(A \circ B\) denotes the element-wise product of \(A\) and \(B\). Furthermore, the set of \(n \times n\)-dimensional symmetric matrices is denoted as \(\mathbb{S}^n\). Besides, \(e_n\) denotes the \(n\)-dimensional vector where all elements equal 1 and \(I_n\) denotes the \(n\)-dimensional identity matrix.

2. Classical SVM model with kernel function and new kernel-free QSSVM model. In this section, we first provide a brief review of the standard SVM model with kernel function. A training data set \(T = \{(x^1, y_1), \ldots, (x^n, y_n)\}\) is given, where \((x^i, y_i) \in \mathbb{R}^m \times \{-1, 1\}\) for \(i = 1, \ldots, n\). Note that \(x^i\) is the explanatory variable for the \(i_{th}\) insurance firm which includes \(m\) different indexes; \(y_i \in \{-1, 1\}\) is an associated label which indicates the firm as failed or healthy respectively. Since the training set in real life can hardly be separated by a linear hyperplane in the original space \((\mathbb{R}^m)\), researchers propose a commonly used method to handle the nonlinearity in the data structure. That is using a nonlinear function \(\phi(x) : \mathbb{R}^m \rightarrow \mathbb{R}^d\) to map the training points to a higher dimensional feature space \((\mathbb{R}^d, d > m)\). Then the SVM model tries to find a hyperplane \(w^T \phi(x) + b = 0\) in the higher dimensional space to satisfy that \(y_i(w^T \phi(x^i) + b) \geq 1\). Note that the margin between two supporting hyperplanes \(w^T \phi(x) + b = -1\) and \(w^T \phi(x) + b = 1\) is denoted by \(\frac{2}{\|w\|}\). And remind that, SVM model aims to separate the training points into two classes while searching the maximal margin approach. Thus, the larger margin indicates the smaller \(\|w\|\). Besides, in order to overcome the linearly inseparable situation, a slack variable vector \(\xi = (\xi_1, \ldots, \xi_n) \in \mathbb{R}^n\) is used to relax the supporting boundary by introducing the misclassification error for each point. Consequently, let \(M > 0\) be the penalty value for the misclassification error, then the soft-margin SVM model can be written as follows.

\[
\begin{align*}
\min_{\frac{1}{2}\|w\|^2 + M\|\xi\|^2} & \\
\text{s.t.} & \quad y_i(w^T \phi(x^i) + b) \geq 1 - \xi_i, \quad i = 1, \ldots, n. \tag{1}
\end{align*}
\]

Note that the value \(M\) reflects the trade-off between the classification error and margin width.

Since the mapping function \(\phi(x)\) is difficult to be handled directly, researchers focus on the dual problem and introduce the kernel function \(K(x, x') = \phi(x)^T \phi(x') : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}\). Let \(K\) denote the kernel matrix where \(K_{ij} = K(x^i, x^j)\). Note that, there are typically several choices for the kernel matrix, such as linear kernel matrix \(K_{ij} = (x^i)^T x^j\), polynomial kernel matrix (degree \(d\)) \(K_{ij} = (\gamma (x^i)^T x^j + r)^d\) \((\gamma > 0, r > 0)\), radial basis function kernel matrix \(K_{ij} = \exp\{-\gamma \|x^i - x^j\|^2\}\) \((\gamma > 0)\) and sigmoid kernel matrix \(K_{ij} = \tanh(\gamma (x^i)^T x^j + r)\) \((\gamma > 0, r > 0)\). Let \(z \in \mathbb{R}^n\) denote the dual variable vector, \(\lambda = (\frac{1}{\|x^1\|}, \ldots, \frac{1}{\|x^n\|})^T\) and \(\Lambda = (K + \text{Diag}(\lambda)) \circ yy^T\), then the dual problem of problem (1) can be formulated as follows [16].

\[
\begin{align*}
\max & \quad -\frac{1}{2} z^T \Lambda z + c^T z_i \\
\text{s.t.} & \quad y^T z = 0, \\
& \quad z \geq 0^n. \tag{2}
\end{align*}
\]

It is worth pointing out that once the optimal solution \(z^*\) of problem (2) is obtained, the optimal solution \(b^*\) can be calculated by \(b^* = y_j - \sum_{i=1}^{n} y_i z_i^* K(x^i, x^j)\). Then
Then we construct a matrix $W$ of generate a vector $u$ to separate $n$ in the data structure, the FQSSVM model directly generates a quadratic surface the training points into a higher dimensional space to deal with the nonlinearity result in machine learning area [18]. Instead of using a kernel function which maps quadratic surface support vector machine (FQSSVM) model which is a state-of-the-art applicable for the users and may not perform well due to some inappropriate choices.

To resolve these concerns, here we briefly introduce the non-kernel fuzzy quadratic surface support vector machine (FQSSVM) model which is a state-of-the-art result in machine learning area [18]. Instead of using a kernel function which maps the training points into a higher dimensional space to deal with the nonlinearity in the data structure, the FQSSVM model directly generates a quadratic surface to separate $n$ training points into two classes in the original space. Note that the quadratic surface is defined as follows.

$$ f(x) = \text{sign}(\sum_{i=1}^{n} y_i z_i^* K(x_i, x) + b^*), $$

(3)

where sign denotes the sign function which extracts the sign of a real number.

However, there is no general guideline for the users to select a proper kernel function for any given data set. The task of searching such a suitable kernel function is always very formidable and time-consuming [24]. Besides, the performance of a classical SVM model depends heavily on the choice of the kernel function and its corresponding parameter set. Therefore, the traditional SVM models are not very applicable for the users and may not perform well due to some inappropriate choices.

Note that since the data set is very imbalanced (the number of insolvent insurers is much less than that of healthy insurers), each training point should not be indiscriminately treated in the model, otherwise the characteristics of the “minority” may not be recognized. In order to deal with this problem, we can assign different discriminately treated in the model, otherwise the characteristics of the “minority” is much less than that of healthy insurers), each training point should not be in-

It is worth pointing out that we can derive a simplified model by the following steps [18]. First, we use the $\frac{m^2+m}{2}$ elements in the upper triangle part of matrix $A$ to generate a vector $u$ as follows.

$$ u \doteq (A_{11}, A_{12}, ..., A_{1m}, A_{22}, A_{23}, ..., A_{2m}, ..., A_{mm})^T. $$

(6)

Then we construct a matrix $W^i \in \mathbb{R}^{m \times \frac{m^2+m}{2}}$ in the following way. For the $j$th row of $W^i$, $j = 1, 2, ..., m$, if the $k$th element of $u$ is $A_{pj}$ or $A_{jp}$ for some $p = 1, 2, ..., m$, $W^i_{jk} = x^i_p$; Otherwise, $W^i_{jk} = 0$. Afterwards, let

$$ H^i = W^i, I_m \in \mathbb{R}^{m \times \frac{m^2+m}{2}}, i = 1, ..., n, $$

$$ D = \sum_{i=1}^{n} w_i(H^i)^T H^i \in \mathbb{R}^{\frac{m^2+m}{2}}, $$

$$ g(x) \doteq \frac{1}{2} x^T A x + b^T x + c = 0, $$

(4)

where $A \in \mathbb{R}^m$, $b \in \mathbb{R}^m$ and $c \in \mathbb{R}$. It is worth pointing out that there is no restriction on matrix $A$, vector $b$ and constant $c$. Thus, the quadratic surface can be any shape which has a lot of flexibility in classification.

Note that since the data set is very imbalanced (the number of insolvent insurers is much less than that of healthy insurers), each training point should not be indiscriminately treated in the model, otherwise the characteristics of the “minority” may not be recognized. In order to deal with this problem, we can assign different fuzzy weights $w_i$, $i = 1, \cdots, n$ to various training points in the model. Suppose the bad/good ratio among the training points is $\gamma$, then for the good training points, let $w_i = 1$; while for the bad training points, let $w_i = \frac{1}{\gamma}$. For a point $x'$, its relative geometrical margin is measured by the segment which intercepts two surfaces $g(x) = g(x')$ and $g(x) = 0$ along its gradient direction $Ax' + b$. Then follow the similar idea in classical SVM model and combine the fuzzy weights, the soft-margin FQSSVM model can be written as follows.

$$ \min_{\xi} \sum_{i=1}^{n} w_i \|Ax'_i + b\|^2 + M \sum_{i=1}^{n} w_i \xi_i^2 $$

$$ \text{s.t. } y_i (\frac{1}{2}(x'_i)^T Ax'_i + b^T x'_i + c) \geq 1 - \xi_i, \ i = 1, ..., n. $$

(5)
\[ v \triangleq (u^T, b^T)^T \in \mathbb{R}^{\frac{n^2 + 3m}{2}}, \]
\[ z^i \triangleq \left( \frac{1}{2} x_1^i x_1^i, x_1^i x_2^i, \ldots, x_1^i x_m^i, \frac{1}{2} x_2^i x_2^i, x_2^i x_3^i, \ldots, x_2^i x_m^i, \ldots, \frac{1}{2} x_m^i x_m^i \right) \in \mathbb{R}^{\frac{n^2 + 3m}{2}} \]

Then, we can reformulate problem (5) as follows.

\[
\begin{align*}
\min & \quad v^T D v + M \sum_{i=1}^{n} w_i \xi_i^2 \\
\text{s.t.} & \quad y_i (v^T z^i + c) \geq 1 - \xi_i, \quad i = 1, \ldots, n.
\end{align*}
\] (8)

Since \( D \) is positive semi-definite, problem (8) is a convex quadratic programming problem with linear constraints. Therefore, it can be efficiently solved by any convex solver. Note that once we get the optimal solutions \( u^* \) and \( c^* \) of problem (8), we can easily derive the optimal matrix \( A^* \) and vector \( b^* \). Then the final decision function for each new variable \( x \) can be constructed as follows.

\[
g(x) = \text{sign} \left( \frac{1}{2} x^T A^* x + (b^*)^T x + c^* \right). \] (9)

**Remark 1.** The FQSSVM model does not use a quadratic kernel in the model, instead it directly uses a nonlinear (quadratic) surface to classify the data points in the original space. In this way, it avoids the forbidding task for searching a proper kernel function and its corresponding parameters in classical SVM models. Therefore, it can save much more effort of the users and be more applicable for them (even some unexperienced starters). Besides, it is more generalized for all types of data. Moreover, the convex structure of FQSSVM makes it easy to be solved. Thus, it is able to handle some large-sized problems. Furthermore, FQSSVM model yields more accurate classifications than classical SVM models on a majority of data. Above all, the non-kernel FQSSVM model is the most advanced technique in the current research. It is a new trend in the machine learning area and demonstrates a big potential in some real applications.

3. Data. In this paper, the models are tested using samples constituted by both solvent and insolvent nonlife firms from USA. The list of these insurers are collected on ISIS of BvD (Bureau van Dijk) (https://isis.bvdep.com/ip). Note that according to the insurers' states, ISIS has divided the insurers into three categories, active, inactive and unknown. “Active” state indicates that the corresponding firm is still in a good condition at the latest closing date. “Inactive” state means the corresponding firm has been liquidated or bankrupted or dissolved at the closing date of the last available year. “Unknown” implies that the state of corresponding firm is just unknown at the closing date of the last available year.

In the numerical experiment, we use samples from “active” and “inactive” states as the training and testing points. Then after deleting those points with missing information, finally there are 987 nonlife firms in the “active” set and 75 nonlife firms in the “inactive” set, respectively.

In order to improve the prediction accuracy, we try to capture the economic situations in certain years and consider their corresponding effects on the solvency issue of nonlife insurers. Note that the following five factors (GDP growth rate, M2 growth rate, inflation rate, interest rate and unemployment rate) are commonly used to reflect the macroeconomic condition [1]. Therefore, we incorporate these five factors into our model to check whether they are helpful in enhancing the prediction performance. It is worth pointing out that these five factors may not be all closely related to insurers' insolvency. Thus, we need to do the feature
selection to eliminate the unrelated features among them. The information of these macroeconomic factors can be derived from the Economist Intelligence Unit (EIU) Country data (https://eiu.bvdep.com/countrydata).

The list of full firm-level explanatory variables for nonlife insurers is shown in Table 1. For each financial variable, in consideration of the latest closing date, “active” insurers extract the data of the latest three years from 2013 to 2015, while “inactive” insurers draw the data three years prior to their own closing date of the last available year. Similarly, we match the macroeconomic information in certain years with the corresponding “active” and “inactive” firms.

Table 1. The firm-level explanatory variables for non-life insurers

| Index | Figure                          | Index | Ratio                           |
|-------|---------------------------------|-------|---------------------------------|
| F1    | Surplus                         | R1    | Net Premium / Surplus           |
| F2    | Net Technical Reserves          | R2    | Tech. Res. / Net Premium        |
| F3    | Total Other Liabilities         | R3    | Tech. Res. / Surplus            |
| F4    | Total Liabilities               | R4    | Liq. Assets / Tech. Res.+ Oth. Liabs |
| F5    | Total Investments               | R5    | Combined Ratio                  |
| F6    | Total Other Assets              | R6    | Expense Ratio                   |
| F7    | Total Assets                    | R7    | Loss Ratio                      |
| F8    | Gross Premium Written           | R8    | Investment Yield                |
| F9    | Net Premium Written             | R9    | Pre-Tax Profitability           |
| F10   | Net Premium Earned              | R10   | Liq. Assets/Net Tech. Res.      |
| F11   | Underwriting Expenses           | R11   | Liq. Ast.+Debts ced Co/Net Tech. Res.+Oth Liabs |
| F12   | Underwriting Result             | R12   | Profit Bef. Tax / Net Prem. Written |
| F13   | Net Investment Income           | R13   | Gross Premium/Surplus           |
| F14   | Profit Before Tax               | R14   | Change in Surplus               |
| F15   | Syndicate Profit                | R15   | Change in Technical Reserves    |
| F16   | Profit After Tax                | R16   | Change in Net Premiums Written  |
| F17   | Profit After Names Expenses     |       |                                 |

Now the training points with full explanatory variables can be applied to different prediction techniques. However, not all the provided variables are actually useful in the forecast. Some unrelated factors can even worsen the performance [9]. Besides, different orders of variables may bias the model and affects the accuracy of prediction. Therefore, we need to do some feature selection and normalization approach before we do the numerical experiments.

First, we independently normalize each feature component by linearly scaling the original data into the range of (-1,1). This process prevents the smaller value input attributes from being overwhelmed by larger value inputs. Hence, normalization can help to reduce prediction errors [20].

Second, for each feature, we use the independent samples t-test to compare the difference between the means of solvency and insolvency firms. Following the t-test the features which do not show significant differences are deleted from the data set.

Third, we use the weight of evidence (WOE) to calculate the information value (IV) for each variable. Note that for the discrete variable, let $\text{WOE}_i$ denote the WOE value for its $i_{th}$ attribute. Meanwhile, for the continuous variable, we first need to separate its value interval into several small intervals by discretization, then let $\text{WOE}_i$ denote the WOE value for its $i_{th}$ small interval. The equation of WOE is given as follows.

$$\text{WOE}_i = \ln \left( \frac{g_{ib}}{gb_i} \right), \quad (10)$$
where \( g_i \) is the number of insurers with good state which corresponds to the \( i_{th} \) attribute or small interval, \( b_i \) is the number of insurers with bad state which corresponds to the \( i_{th} \) attribute or small interval, \( g \) is the total number of insurers with good state in the sample and \( b \) is the total number of insurers with bad state in the sample. Then we can calculate the information value for each variable by the following equation:

\[
IV = \sum_i \left\{ \left( \frac{g_i}{g} - \frac{b_i}{b} \right) \times \ln \left( \frac{g_i}{g} \div \frac{b_i}{b} \right) \right\} = \sum_i \left\{ \left( \frac{g_i}{g} - \frac{b_i}{b} \right) \times \text{WOE}_i \right\}.
\]  

Note that the bigger the IV is, the more discriminative power the corresponding variable has. As usual, we choose the variables with IV bigger than 0.02 into the final model. It is worth pointing out that, among these 38 variables (5 macroeconomic attributes and 33 original attributes), four variables, namely Combined Ratio (by \( t \)-test), Investment Yield (by \( t \)-test), inflation rate (by \( t \)-test) and M2 growth rate (by IV), are excluded in the final model.

4. Method comparison. Note that predicting financial failure is originally a classification problem to categorize organizations as healthy or non-healthy ones. Many researchers focused on this kind of problem and developed a number of methodologies. In the numerical experiment, we will pick several successful and commonly used benchmark methods in the literature and compare their performances with our new method. This list includes artificial neural network (ANN) [15], traditional support vector machine (SVM) [9] and multivariate statistical methods such as multivariate discriminant analysis (MDA) [14], logistic regression analysis (LRA) [5] and stepwise logistic regression (SLR) [33].

ANN is a parallelized computational model in computer science and other research disciplines, which is able to learn from examples and adapt to new situations. It is based on a large set of simple neural units (artificial neurons), a transfer function, a pattern of connectivity and a propagation and activation rule. ANN is loosely analogous to the observed behavior of a biological brain’s axons. Each neural unit is connected with many others, and links can enhance or inhibit the activation state of adjoining neural units. Each individual neural unit computes using summation function. There may be a threshold function or limiting function on each connection and on the unit itself, such that the signal must surpass the limit before propagating to other neurons. These systems are self-learning and trained, rather than explicitly programmed. ANN typically consist of multiple layers, and the signal path traverses from the input layer to the output layer. In this process, back propagation uses forward stimulation to reset weights on the neural units. Generally, ANN includes two working phases, named learning and recall. During the learning phase, known data sets are commonly used as a training signal. Then the recall phase uses the weight obtained in the learning phase to get the final result. Following convention methods, we use the classical three hidden layers architecture in ANN and let the number of neurons in the input layer equal to the dimension of the data set in our experiment. The ultimate functional relationships between the input units and the output prediction units are based on a linear aggregation function used in conjunction with a logistic activation function. Note that the numbers of neurons in the remaining two layers are determined in the following way. The corresponding layer starts as an empty one and new neurons are added to it at a time. This process stops when the new added neurons cannot affect the network any more. Since the output signal is a binary variables (0 indicates “inactive” while 1
indicates “active”), we also use the sigmoid activation function in the output layer. It is worth pointing out that ANN is performed on MATLAB in our experiment.

Traditional SVM is an important classification and pattern recognition technique. It follows the structural risk minimization (SRM) idea and uses the maximal margin approach to find a hyperplane that separates the training points into two classes. For the nonlinear structural data, traditional SVM maps original space into a high-dimensional feature space, then finds an optimal hyperplane which maximizes the margin between itself and the nearest training examples in the new high-dimensional space and minimizes the expected classification error. It is worth pointing out that the mapping function and corresponding kernel matrix decide the complexity of the classification and largely affect the performance of the traditional SVM model. In the numerical experiment, we choose the kernel matrix between Gaussian kernel matrix and linear kernel matrix which are the top two commonly used ones in literature. The corresponding parameters are searched by the grid method.

For the multivariate statistical methods, we employ three classical ones, namely multivariate discriminant analysis (MDA), logistic regression analysis (LRA) and stepwise logistic regression (SLR) to predict the insolvency of non-life insurers. MDA is a method which compresses a combination of predictor variables by a linear discriminant function to yield a discriminant score. Then based on this discriminant score, a new and unlabeled observation can be classified into its appropriate group. Note that the corresponding linear discriminant function can be written as follows.

\[ S = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n, \]  

(12)

where \( S \) is a score, \( a_0 \) is an estimated constant value, \( a_i \) are estimated coefficients for predictor variables \( x_i \), \( i = 1, \ldots, n \). In the numerical experiment, MDA is first applied to the training data to characterize the group differences, then to classify the unknown insurers. Comparatively, LRA is a form of regression in which the dependent variables are usually dichotomies while the independent variables are of any type. In the numerical experiment, the dependent variable takes the value 1 (solvency) with a probability of success \( p(z_i) \), or the value 0 (insolvency) with probability \( 1 - p(z_i) \). Here the cumulative probability function \( p(z_i) \) between the predictor independent variables and binary dependent is expressed as follows.

\[ p(z_i) = \frac{1}{1 + e^{-z_i}}, \quad z_i = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n, \]  

(13)

where \( b_0 \) is an estimated constant value, and \( b_m \) are estimated coefficient for predictor variables \( x_i \), \( i = 1, \ldots, n \). Furthermore, SLR is an extension of LRA in which all the variables are not simultaneously introduced into the model. Instead, it ranks the variables from high level to low level according to their influences on the dependent variable. Then SLR follows this sequence to add the variables to the logistic regression model step by step.

5. **Numerical tests.** In this section, we conducted several numerical experiments to investigate the performance of different methods in predicting insolvency. Moreover, in order to show the effect of macroeconomic factors and feature selection in improving the predictive accuracy, we show the results with and without this information for each model. Here, \( C_1 \) denotes the case where no macroeconomic factor is added and no feature selection is implemented; \( C_2 \) denotes the case where
Table 2. A confusion matrix

| Prediction | Good | Bad |
|------------|------|-----|
| Good       | Correct goods (true negative) | Type II error (false negative) | Goods predicted |
| Bad        | Type I error (false positive) | Correct bads (true positive) | Bads predicted |
| Total      | Goods observed | Bads observed | Sample size |

Table 3. Overall accuracy results for different methods in various cases

| Case | Methods     |
|------|-------------|
|      | ANN | SVM | MDA | LRA | SLR | FQSSVM |
| C₁   | 0.816 | 0.861 | 0.828 | 0.831 | 0.833 | 0.865 |
| C₂   | 0.837 | 0.881 | 0.862 | 0.864 | 0.850 | 0.888 |
| C₃   | 0.862 | 0.903 | 0.840 | 0.848 | 0.851 | 0.915 |

Macroeconomic factors are added but no feature selection is implemented; C₃ denotes the case where both macroeconomic factors and feature selection are used in the models.

In each case, we repeated the tests for 50 times. And to avoid being affected by the biased learning and testing samples, we applied the 10-cross-validation in each test. Note that ten sets were randomly generated each time. In the ith experiment, the ith set was regarded as the test set while the other nine sets became the training sets. After all tests in each case, every result would be the average value.

Following the traditional way, the performance of different methods was measured by the predicative accuracy. Note that the confusion matrix in Table 2 is a good tool to gauge the prediction accuracy or level of misclassification [27]. The overall accuracy of the model is the percentage of the correctly predicted points in the whole testing points. Moreover, we also provide the type I and type II errors to show more detailed information about the performance. Here, a type I error denotes that a good company is predicted as a bad one (false positive), while a type II error denotes that a bad company is predicted as a good one (false negative).

Furthermore, the receiver operation curve (ROC) also provides a good measure for the discriminant power. It is a graph which plots the true positive rate (sensitivity) against the false positive rate (fall-out) at various threshold settings. Note that if two classes can be perfectly classified by a model, the ROC curve would go along the edges of the square, whereas if a model performs just like a random guess, the curve would go along the diagonal line. Thus, the area under ROC (AUC) which ranges between 0.5 and 1 is an indicator of how good the corresponding model is.

Besides, since the efficiency of solving model is tremendously important for some real applications, computational time is another significant measurement of model performance. Particularly, an efficient method can deal with the big-sized problem which is increasingly prevalent in the current age of big data.

For the implementation issue, our FQSSVM model is solved by the well-known solver CVX [13]. The grid searching method is used to get the optimal penalty parameter \( M \) i.e. \( \log_2 M \in \{-9, -8, ..., 0, ..., 9\} \). Note that the accuracy criterion \( \epsilon \) is set as \( 10^{-3} \) for all models. All the computational experiments in this paper are carried out by matlab 7.9.0 on a PC equipped with 3.3 Ghz Intel Core i5 CPU and 8G usable RAM.

The overall accuracy results for different methods in various cases are presented in Table 3. The model with the best accuracy is highlighted in bold. It is obvious that
Table 4. Type I errors for different methods in various cases

| Case | Type I error | ANN | SVM | MDA | LRA | SLR | FQSSVM |
|------|--------------|-----|-----|-----|-----|-----|--------|
| C₁   | 0.185        | 0.141| 0.173| 0.170| 0.168| 0.137|        |
| C₂   | 0.165        | 0.121| 0.139| 0.137| 0.151| 0.115|        |
| C₃   | 0.139        | 0.099| 0.161| 0.154| 0.151| 0.088|        |

Table 5. Type II errors for different methods in various cases

| Case | Type II error | ANN | SVM | MDA | LRA | SLR | FQSSVM |
|------|---------------|-----|-----|-----|-----|-----|--------|
| C₁   | 0.174         | 0.117| 0.159| 0.155| 0.153| 0.108|        |
| C₂   | 0.135         | 0.094| 0.128| 0.119| 0.136| 0.079|        |
| C₃   | 0.127         | 0.068| 0.146| 0.130| 0.128| 0.047|        |

Table 6. AUC results for different methods in different comparisons

| Case | Methods       | ANN | SVM | MDA | LRA | SLR | FQSSVM |
|------|---------------|-----|-----|-----|-----|-----|--------|
| C₁   | 0.832         | 0.861| 0.839| 0.835| 0.829| 0.856|        |
| C₂   | 0.843         | 0.873| 0.841| 0.853| 0.862| 0.887|        |
| C₃   | 0.840         | 0.882| 0.841| 0.851| 0.856| 0.892|        |

Table 7. Computational times (in seconds) for different methods in different comparisons

| Case | Methods       | ANN | SVM | MDA | LRA | SLR | FQSSVM |
|------|---------------|-----|-----|-----|-----|-----|--------|
| C₁   | 201.6         | 347.9| 37.17| 40.05| 213.8| 110.3|        |
| C₂   | 214.3         | 336.2| 36.91| 39.28| 195.1| 98.65|        |
| C₃   | 196.8         | 329.5| 36.85| 39.44| 187.6| 91.41|        |

Table 8. Main pros and cons of different methods

| Method  | Pros                              | Cons                                |
|---------|-----------------------------------|-------------------------------------|
| ANN     | handle nonlinear structure        | black-box character                 |
|         | adaptability to environments      | difficult to interpret              |
|         | resistance to noise               | hard to analyze results             |
|         | pattern recognition               | parameters choice                   |
| SVM     | handle nonlinear structure        | kernel function choice              |
|         | distribution-free                 | parameters choice                   |
|         | good performance                  | computational times                 |
| MDA     | easy to implement                 | can’t handle nonlinear structure    |
|         | computational times               | strong assumption on data           |
| LRA     | easy to implement                 | can’t handle nonlinear structure    |
|         | computational times               | strong assumption on data           |
| SLR     | easy to implement a proper subset | can’t handle nonlinear structure    |
|         | of variables                      | strong assumption on data           |
|         | computational times               | computational times                 |
| FQSSVM  | handle nonlinear structure        | dimension expansion                 |
|         | no need for kernel function       |                                     |
|         | distribution-free                 |                                     |
|         | good performance                  |                                     |
|         | good generalization               |                                     |
the SVM models (kernel-SVM and FQSSVM) beat other machining learning and statistical techniques in all cases. Averagely, they increase the predictive accuracy by 3~4 percentage points. Moreover, between the SVM models, FQSSVM slightly outperforms kernel-SVM. Therefore, these results strongly demonstrate the effective power of our new method in predicting nonlife insurer’s insolvency. Furthermore, the detailed information about Type I and Type II errors are shown in Table 4 and Table 5, respectively. Note that the SVM models lead other techniques in reducing both Type I and Type II errors. Since the loss of a missed bad insurer is much bigger than the loss of a misclassified good insurer, Type II error is much more important than Type I error in predicting insolvency. The error results indicate that our method indeed did the best job in controlling Type II error while keeping Type I error at a very low level. Thus, the proposed method is more practical than other benchmark methods in some real applications. Moreover, we also calculate the average values of AUC to check the discriminant power among these methods. The corresponding results are summarized in Table 6. As noticed, the SVM models are evident in improving AUC among these methods. Besides, for each case, we randomly extract the result of one test from the repeated sampling/modelling. Then the ROC curves are drawn for Case 1, Case 2 and Case 3 in Fig 1, Fig 2 and Fig 3, respectively. Overall, the curves of FQSSVM are generally above the curves of other methods which indicates that FQSSVM has the best discriminant power.

![Figure 1. ROC curves for one test in case C_1](image)

Furthermore, we compare the efficiencies of these methods. The computational times (in seconds) are presented in Table 7. It is worth pointing out that the searching time for a proper kernel function is included in the kernel-SVM method. From the results, we can see that the efficiency of FQSSVM is consistently much better than ANN, kernel-SVM and SLR in all cases. That is because there is no need for a FQSSVM model to choose a kernel function and corresponding parameters. Therefore, FQSSVM saves a lot of time in searching the optimal setting for a certain data set. Moreover, the convex structure of FQSSVM guarantees its high efficiency in solving process. On the other hand, although MDA and LRA have relatively shorter computational times, considering the performance, FQSSVM is a much better choice than them in practice.
At last, we summarize the pros and cons of each method in Table 8. Note that compared with the existing benchmark methods, SVM models show their superiority in predicting insolvency. Specially, in contrast with the traditional kernel-SVM model, our proposed method based on the state-of-the-art FQSSVM model is more universal for all kinds of problems and much easier to be solved. On the other hand, it is worth pointing out that the downside of our FQSSVM model is the dimension expansion in the corresponding optimization problem (the original $m$-dimensional problem should be lifted to an $\frac{m^2+3m}{2}$-dimensional problem). Thus, if there exist thousands of variables in the model, the problem size may become an issue and increase the computational time. However, so many variables are seldom seen in any model simultaneously, especially combined with the feature selection process. Hence, FQSSVM model should be more suitable for the users and has a bigger potential in real practice.
6. Conclusion. In this paper, we have proposed a new prediction method, FQSSVM, which is based on a non-kernel fuzzy quadratic surface SVM to predict the insolvency of nonlife insurers. Unlike the traditional way which only focuses on the firm-level financial ratios of nonlife insurance companies, we incorporates the macroeconomic factors to provide a more comprehensive and accurate model. Besides, we also apply feature selection to exclude some unrelated variables which may affect the accuracy of the classification. Note that compared with other benchmark methods and traditional SVM models which are commonly used in predicting solvency, our method is more general and doesn’t need to search for a proper kernel function and corresponding optimal parameters. More important, our model is the most effective one (in terms of a reasonable computational time) and generates the most accurate prediction results. Specifically, comparing to other methods, our model has the most accurate prediction rate and least Type I and Type II errors. The promising numerical results based on the real data strongly demonstrate the superiority of the proposed method and indicate its big potential.

Acknowledgment. Tian’s research has been supported by the National Natural Science Foundation of China Grants #11401485 and # 71331004.

REFERENCES

[1] M. Al-Smadi, Credit Risk, Macroeconomic and Bank Specific Factors, VDM Verlag Dr. Müller, 2011.
[2] E. Baranoff, T. Sager and T. Shively, A semiparametric stochastic spline model as a managerial tool for potential insolvency, Journal of Risk and Insurance, 67 (2000), 369–396.
[3] A. Best, Best’s insolvency study-property/casualty insurers, Best’s Review-Property/Casualty Insurance Edition, (1991), 16–23.
[4] J. Carson and R. Hoyt, Life insurer financial distress: Classification models and empirical evidence, Journal of Risk and Insurance, 62 (1995), 764–775.
[5] J. Cheng and M. Weiss, The role of rbc, hurricane exposure, bond portfolio duration, and macroeconomic and industry-wide factors in property-liability insolvency prediction, Journal of Risk and Insurance, 79 (2012), 723–750.
[6] H. Chew and C. Lim, On regularisation parameter transformation of support vector machines, Journal of Industrial and Management Optimization, 5 (2009), 403–415.
[7] S. Cho, J. Kim and J. Bae, An integrative model with subject weight based on neural network learning for bankruptcy prediction, Expert Systems with Applications, 36 (2009), 403–410.
[8] J. Cummins, M. Grace and R. Phillips, Regulatory solvency prediction in property-liability insurance: Risk-based capital, audit ratios, and cash flow simulation, Journal of Risk and Insurance, 66 (1999), 417–458.
[9] U. Dellepiane, M. Marcantonio, E. Laghi and S. Renzi, Bankruptcy prediction using support vector machines and feature selection during the recent financial crisis, International Journal of Economics and Finance, 7 (2015), 182–194.
[10] A. Dimitras, R. Slowinski, R. Susmaga and C. Zopounidis, Business failure using rough set, European Journal of Operational Research, 114 (1998), 263–280.
[11] P. Du Jardin, Predicting bankruptcy using neural networks and other classification methods: The influence of variable selection techniques on model accuracy, Neurocomputing, 73 (2010), 2047–2060.
[12] M. Grace, S. Harrington and R. Klein, Risk-based capital and solvency screening in property-liability insurance: Hypothesis and empirical tests, Journal of Risk and Insurance, 65 (1998), 213–243.
[13] M. Grant and S. Boyd, Cvx: Matlab Software for Disciplined Programming, version 1.2, Technical report, http://cvxr.com/cvx 2010.
[14] S. Hsiao and T. Whang, A study of financial insolvency prediction model for life insurers, Expert Systems with Applications, 36 (2009), 6100–6107.
[15] M. Kim and D. Kang, Ensemble with neural networks for bankruptcy prediction, Expert Systems with Applications, 37 (2010), 3373–3379.
[16] G. Lanckriet, N. Cristianini, P. Bartlett, L. El Ghaoui and M. Jordan, Learning the kernel matrix with semi-definite programming, *Journal of Machine Learning Research*, 5 (2004), 27–72.

[17] S. Lee and J. Urrutia, Analysis and prediction of insolvency in the property-liability insurance industry: A comparison of logit and hazard models, *Journal of Risk and Insurance*, 63 (1996), 121–130.

[18] J. Luo, S.-C. Fang, Y. Bai and Z. Deng, Fuzzy quadratic surface support vector machine based on fisher discriminant analysis, *Journal of Industrial and Management Optimization*, 12 (2016), 357–373.

[19] J. Maria, S. Sancho and B. Carlos, Prediction of insolvency in non-life insurance companies using support vector machines, genetic algorithms and simulated annealing, *Fuzzy Economic Review*, 9 (2004), 79–94.

[20] J. Min and Y. Lee, Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters, *Expert Systems with Applications*, 28 (2005), 603–614.

[21] S. Min, J. Lee and I. Han, Hybrid genetic algorithms and support vector machines for bankruptcy prediction, *Expert Systems with Applications*, 31 (2006), 652–660.

[22] S. Pottier and D. Sommer, Empirical evidence on the value of group-level financial information in insurer solvency surveillance, *Risk Management and Insurance Review*, 14 (2011), 73–88.

[23] S. Sancho, D. Mario, J. Maria, P. Fernando and B. Carlos, Feature selection methods involving support vector machines for prediction of insolvency in non-life insurance companies, *Intelligent Systems in Accounting, Finance and Management*, 12 (2004), 261–281.

[24] K. Schittkowski, Optimal parameter selection in support vector machines, *Journal of Industrial and Management Optimization*, 1 (2005), 465–476.

[25] B. Scholkopf and A. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, 2002.

[26] K. Shin, T. Lee and H. Kim, An application of support vector machines in bankruptcy prediction model, *Expert Systems with Applications*, 28 (2005), 127–135.

[27] N. Siddiqi, *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*, John Wiley & Sons, 2015.

[28] Y. Tian, M. Sun, Z. Deng, J. Luo and Y. Li, A new fuzzy set and non-kernel SVM approach for mislabeled binary classification with applications, *IEEE Transactions on Fuzzy Systems*, 25 (2017), 1536–1545.

[29] V. Vapnik, *Statistical Learning Theory*, John Wiley & Sons, Inc., New York, 1998.

[30] C. Wu, G. Tzeng, Y. Goo and W. Fang, A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy, *Expert Systems with Applications*, 32 (2007), 397–408.

[31] C. Xie, C. Luo and X. Yu, Financial distress prediction based on SVM and mda methods: The case of Chinese listed companies, *Quality & Quantity*, 45 (2011), 671–686.

[32] Z. Yang, W. You and G. Ji, Using partial least squares and support vector machines for bankruptcy prediction, *Expert Systems with Applications*, 38 (2011), 8336–8342.

[33] L. Zhang and N. Nielson, Solvency analysis and prediction in property casualty insurance: Incorporating economic and market predictors, *Journal of Risk and Insurance*, 82 (2015), 97–124.

Received November 2017; 1st revision January 2018; 2nd revision January 2018.

E-mail address: yetian@swufe.edu.cn
E-mail address: yangwu@swufe.edu.cn
E-mail address: glai@uncc.edu
E-mail address: 757603259@qq.com