Performance evaluation of least-squares probabilistic classifier for corporate credit rating classification problem

Miho Saito\(^1\) and Suguru Yamanaka\(^2\)∗

\(^1\) Graduate school of science and engineering, Aoyama Gakuin University, 5-10-1 Fuchinobe, Chuo-ku, Sagamihara-shi, Kanagawa 232-5258, Japan
\(^2\) College of science and engineering, Aoyama Gakuin University, 5-10-1 Fuchinobe, Chuo-ku, Sagamihara-shi, Kanagawa 232-5258, Japan

∗Corresponding author: syamanaka@gem.aoyama.ac.jp

Received September 16, 2020, Accepted January 6, 2021

Abstract

The corporate credit rating classification problem has attracted lots of research interests in the literature of financial risk management. This article introduces the least-squares probabilistic classifier to the problem in an attempt to provide a model with better explanatory power. Empirical results show that the least-squares probabilistic classifier outperforms the logistic regression model, random forest, and the support vector machine in prediction accuracy ratios and F1 scores, for the samples of bond issuer firms in Japan.

Keywords credit risk, credit rating, machine learning, multi-category classification, least-squares probabilistic classifier

Research Activity Group Mathematical Finance

1. Introduction

Corporate credit ratings have been extensively used by banks and bond investors as a surrogate measure of default risk of debtors. Credit ratings are one of the determinants of lending interest rates and even the liquidity of bonds. Banks and credit rating agencies have been developing reliable quantitative methods for automatic classification systems for credit rating classification problems.

Many researchers have attempted to construct automatic classification systems to solve credit rating problems using traditional statistical methods and machine learning techniques. The former include linear regression [1–3], probit regression [4, 5], logistic regression [6–8]. The latter consists of neural networks [9–11], random forests [12] and support vector machines (SVM) [13–17]. In particular, logistic regression, random forests, and SVM are the method most commonly used in financial practice.

We conduct an empirical analysis to assess the usefulness of the least-squares probabilistic classifier (LSPC) for credit rating classification, with Japanese borrower firms as default samples. LSPC is a multi-class probabilistic classification method proposed in [18], which is computationally very efficient and numerically stable. We present an empirical evaluation of LSPC, logistic regression, random forests, and SVM in order to assess their effectiveness for credit rating prediction.

The remainder of this paper is organized as follows. Section 2 briefly reviews the formulation of credit rating classification problem and LSPC. Section 3 shows the empirical results on credit rating classification using LSPC. Section 4 concludes.
where $k(x, x')$ are the kernel functions. In this paper, we employ Gaussian kernel $k(x, x') = \exp(-\|x - x'\|^2/2\sigma^2)$ for kernel functions. To estimate parameter $\alpha_y = (\alpha_{y,1}, \alpha_{y,2}, \ldots, \alpha_{y,N})$, we minimize the squared error with regularization term $J_y(\alpha_y) + \lambda\|\alpha_y\|^2$, where

$$J_y(\alpha_y) := \frac{1}{2} \int (q(y|x, \alpha) - p(y|x))^2 dx.$$ 

The estimated parameter can be obtained by solving the following system of linear equations (see [18] for details):

$$(K + \lambda E)\alpha_y = b_y$$

where $K = (K_{ij}) = [(1/N) \sum_{i=1}^{N} k(x_n, x_i)k(x_n, x_j)]$ and $b_y = (b_{y,i}) = [(1/N) \sum k(x_n, x_i)]$. Then, we classify test sample $x$ to class $\hat{y}$ by $\hat{y} = \arg\max_y \{p(y|x)\}$.

To better understanding of LSPC, we show a sample of the image of generated conditional probability by LSPC, obtained in our empirical analysis. Fig. 1 shows the estimated conditional probability of credit rating class (0, 1, 2, 3) over the ratio of non-operating income to net sales (labeled by $x_{i,15}$) and ratio of cash and deposits to net sales (labeled by $x_{i,25}$).

3. Empirical analysis

3.1 Data

To validate the performance of LSPC to credit rating classification, we applied LSPC to a real-world case of credit rating in Japan. Our application is a bond-issuer rating announced by R&I, one of the big credit rating agencies in Japan. We obtained the bond-ratings samples from 1999 to 2019. In our analysis, the learning samples are the credit rating samples observed from 1999 to 2018, and the validation samples are the credit rating samples observed in 2019. The bond-issuer-rating announced by R&I are 9 ranks: AAA, AA, A, BBB, BB, B, CCC, CC, and D. However, we adjusted our data to four classes by combining AAA and AA into one group, and BB, B, CCC, CC, and D into one group, because the respective numbers of companies were small under the 9 classes. The original data consisted of 109 financial-ratio variables that were known to affect credit ratings, as documented in the existing literature.

3.2 Data preprocessing

Data preprocessing is an integral step in machine learning as data quality and useful information it generates directly affect our model’s ability to learn. For the first step in data preprocessing, we handled Null or NaN values of covariates. We handle the problem by the k-NN imputation method proposed in [19].

Next, we employ min-max scaling, which rescales the range of features to the range in [0, 1], in order to mitigate the size effect of variables. Then, we handle the problem of multicollinearity problem. Multicollinearity occurs when predictor variables have high correlations, leading to unreliable and unstable regression coefficient estimates. We employ widely used multicollinearity diagnostic, that is, the variance inflation factor (VIF). VIF for a pair of covariates $x_i$ and $x_j$ is calculated by $VIF = 1/(1 - \text{cor}(x_i, x_j)^2)$, where $\text{cor}(x_i, x_j)$ is the correlation coefficient. In our empirical analysis, concerns arise when VIF is greater than 2.78, which corresponds to a correlation of 0.8 with other variables. If the value of VIF of covariates $x_i$ and $x_j$ overs 2.78, we remove one of them from the candidate list of covariates.

3.3 Variable selection and standardizing variables

We employ elastic-net, which is proposed in [20] to identify candidate covariates, and select 26 variables. Then, we transform covariates such that it can be recognized samples obtained from the standard normal distribution, with a mean of 0 and a standard deviation of 1. Specifically, we employ the Yeo-Johnson transformation.
Table 1. Classification results by LSPC (Upper: Learning samples, Lower: validation sample).

|                | Actual class |
|----------------|--------------|
|                | 0 1 2 3      |
| LSPC: Predicted| 0 637 4 0 0  |
| Class          | 2 0 0 780 0  |
|                | 3 0 0 0 105  |

Table 2. Classification results by logistic regression, random forest, and SVM (Upper: Learning samples, Lower: validation sample, respectively).

|                | Actual class |
|----------------|--------------|
|                | 0 1 2 3      |
| Logistic regression: Predicted | 0 431 192 18 0  |
| Class          | 116 807 211 1 |
|                | 21 226 522 11 |
|                | 3 0 15 45 45  |

Random forest:

|                | Actual class |
|----------------|--------------|
|                | 0 1 2 3      |
|                | 0 24 2 1 0   |
| Predicted      | 1 6 65 18 0  |
| Class          | 2 0 2 31 1   |
|                | 3 0 0 0 1    |

SVM:

|                | Actual class |
|----------------|--------------|
|                | 0 1 2 3      |
|                | 0 622 19 0   |
| Predicted      | 12 1109 14 0 |
| Class          | 20 21 759 0  |
|                | 3 0 1 103    |

Table 3. Accuracy ratio and F1-score on validation samples.

|                | Accuracy ratio | F1-score |
|----------------|----------------|----------|
| LSPC           | 0.801          | 0.806    |
| Logistic regression | 0.623          | 0.629    |
| Random forest  | 0.642          | 0.644    |
| SVM            | 0.715          | 0.725    |

3.4 Results of classification

We implemented LSPC with Python. We set the values of the hyperparameter of LSPC by $\lambda = 0.01$ and $h = 0.25$, referring to the results of our pre-analysis. For the implementation of benchmark models of logistic regression, random forests, and SVM, we employed the machine learning package “scikit-learn”.

Table 1 shows the result of classification by LSPC which is showed in the form of a confusion matrix. We recognize the diagonal components of Table 1 make up a large share of classification results, that implies LSPC solved the classification problem successfully.

In order to evaluate the accuracy of classification results, we introduce a widely used measure called accuracy ratio and weighted F1-scores. The accuracy ratio is defined as the number of correct predictions over the number of predictions. The value of the weighted F1-score is the weighted average of F1-scores for each class label, using the number of true instances for each class label for weights. F1-scores are the harmonic mean of precision and recall, that is, $F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$, where precision is defined as the number of true positives over the number of true positives plus the number of false positives. Recall is defined as the number of true positives over the number of true positives plus the number of false negatives. The calculated accuracy ratio and weighted F1-score for the classification result by LSPC are 0.801 and 0.806 respectively.

To validate the usefulness of our model as a tool for credit rating classification, we also experimented additional three statistical or machine learning techniques that had been adopted in prior studies on credit rating classification: logistic regression, random forest, and support vector machine. Table 2 shows the result of classification by logistic regression, random forest, and SVM, which is shown in the form of a confusion matrix. Table 3 shows comparisons of experimental results. Clearly, LSPC outperforms the other classification methods in accuracy ratio and weighted F1 scores.

In order to find the relative importance of the predictive variables for credit rating classification with LSPC, we employ the technique of the permutation feature importance. The permutation feature importance is defined to be the decrease in a model score when a single feature value is randomly shuffled. Table 4 shows the ranking of the variable’s importance measured by permutation feature importance. The table shows the variables which are categorized in the group of efficiency and safety are relatively important for the classification.
To improve the model prediction under fluctuated business situations, adopting some macro-economic factors for predictive variables is recommended, which is the idea originally suggested in [22] and improved the prediction accuracy of their model. This point will be implemented in the future.

4. Concluding remarks

We have described the modeling procedures for credit rating classification based on LSPC. There were numerous candidate predictive variables that can be considered, therefore, finding the most important predictive variable is very crucial. We selected variables automatically according to the result of elastic net regression, and evaluated variable importance by permutation importance technique. The empirical results show that our model’s information efficiency is superior to the well-known logistic regression model, random forests, and support vector machines.

To improve the model prediction under fluctuated business situations, adopting some macro-economic factors for predictive variables is recommended, which is the idea originally suggested in [22] and improved the prediction accuracy of their model. This point will be implemented in the future.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP18K12818.

References

[1] J. O. Horrigan, The determination of long-term credit standing with financial ratios, J. Account. Res., 4 (1966), 44–62.
[2] T. F. Pogue and R. M. Soldosky, What’s in a bond rating, J. Financial Quant. Anal., 4 (1969), 201–228.
[3] R. R. West, An alternative approach to predicting corporate bond ratings, J. Account. Res., 8 (1970), 118–125.
[4] R. S. Kaplan and G. Urwitz, Statistical models of bond ratings: a methodological inquiry, J. Bus., 52 (1979), 231–261.
[5] R. C. Hwang, K. F. Cheng and C. F. Lee, On multiple-class prediction of issuer credit ratings, Appl. Stochastic Models Bus. Ind., 25 (2009), 535–550.
[6] W. P. H. Poon, M. Firth and H.-G. Fung, A multivariate analysis of the determinants of Moody’s bank financial strength ratings, J. Int. Financial Mark. Inst. Money, 9 (1999), 267–283.
[7] M. Kamstra, P. Kennedy and T.-K. Suan, Combining bond rating forecasts using logit, Financial Rev., 36 (2001), 75–96.
[8] T. Yasukawa, Equal slope assumption and bond rating data: analysis of using ordered logit model and continuation ratio model (in Japanese), in: Proc. Inst. Statist. Math., 50 (2002), 201–216.
[9] J. W. Kim, H. R. Weistroffer and R. T. Redmond, Expert systems for bond rating: a comparative analysis of statistical, rule-based and neural network systems, Expert Syst., 10 (1993), 167–172.
[10] C.-F. Tsai and J.-W Wu, Using neural network ensembles for bankruptcy prediction and credit scoring, Expert Syst. Appl., 34 (2008), 2639–2648.
[11] P. Hajek, Municipal credit rating modelling by neural networks, Decis. Support Syst., 51 (2011), 108–118.
[12] M. Ryser and S. Denzler, Selecting credit rating models: a cross-validation-based comparison of discriminatory power, Financial Mark. Portf. Manag., 23 (2009), 187–203.
[13] Z. Huang, H. Chen, C.-J. Hsu, W.-H. Chen and S. Wu, Credit rating analysis with support vector machines and neural networks: a market comparative study, Decis. Support Syst., 37 (2004), 543–558.
[14] L. Cao, L. K. Guan and Z. Jingqing, Bond rating using support vector machine, Intell. Data Anal., 10 (2006), 285–296.
[15] Y.-C Lee, Application of support vector machines to corporate credit rating prediction, Expert Syst. Appl., 33 (2007), 67–74.
[16] K. Kim and H. Ahn, A corporate credit rating model using multi-class support vector machines with an ordinal pairwise partitioning approach, Comput. Oper. Res., 39 (2012), 1800–1811.
[17] K. Tanaka and H. Nakagawa, A method of corporate credit rating classification based on support vector machine and its validation in comparison of sequential logit model (in Japanese), Trans. Oper. Res. Soc. Jpn., 57 (2014), 92–111.
[18] M. Sugiyama, Superfast-trainable multi-class probabilistic classifier by least-squares posterior fitting, IEICE Trans. Inf. & Syst., E93-D (2010), 2690–2701.
[19] J. Takahashi and T. Yamashita, Missing value imputation for large financial statement database using k-NN method (in Japanese), in: Proc. Inst. Statist. Math., 50 (2002), 954–959.
[20] T. Kaneko and H. Nakagawa, An application of top-down approach for credit loan portfolio based on ratings process (in Japanese), Kinyu Kenkyu, 29 (2010), 19–44.