RMSC4102 GROUP PROJECT
2016-17 SPRING
CHINESE UNIVERSITY OF HONG KONG

Predicting Default Risk with Support Vector Machines

CHU Jing	1155047032
XIE Peijie	1155047076
ZHANG Mengqi	1155047022
ZHOU Zehui	1155046910

Abstract

Predicting default risk is crucial for firms to operate successfully and for banks to assess corporate risk premium. Among various binary classification methods used to predict default risk, SVM is superior due to its flexibility in linearly non-separable situation. This research implements SVM to predict default risk of Polish firms based on the method proposed by Chen, Härdle and Moro (2011). Further improvement is proposed considering imbalanced class proportion and class-dependent misclassification costs. The improved model with the objection to minimize the expected loss is more practical for investors.

Table of Contents

1. Introduction

Default risk has become a heated topic recently as a default storm is brewing over china bond market. Modeling default probabilities helps firms to deduce credit cost and investors to make wise decisions. From the perspective of risk management, it is therefore of crucial importance to choose a well-performed predictive model. Historically, a large variety of statistical models were proposed, including discriminant analysis (DA) (Beaver, 1966), logistic analysis (Ohlson, 1980), Artificial Neural Networks (ANN) (Hertz, Krogh & Palmer, 1991) and Support Vector Machines (SVM). In this study, we predict bankruptcy probabilities of Polish firms using financial ratios based on the SVM method introduced in *Modeling default risk with support vector machines* by Chen, Härdle and Moro (2011). The major novelty of our study is to incorporate class-dependent misclassification costs and derive a more practical model. The report is organized as follows. Section 2 introduces the SVM classifying method. Section 3 reviews relevant literature, particularly summarizing the study of Chen, Härdle and Moro (2011), on which our research is based. In Section 4 and 5, our methodology and results are presented. Finally, Section 6 concludes our research and discusses some possible extensions.

2. The Support Vector Machine

The Support Vector Machine is a binary classification method developed by Vapnik and his colleagues at Bell laboratories (Vapnik, V., 1995). For a binary problem, a training dataset $\{x_i, y_i\}$, i = 1, 2, …, n, is given, where $x_i \in R^d$ and $y_i = \{+1, -1\}$. The objection is to find a separating hyperplane with classifying function satisfying $f(x) \geq 0$ for $y_i = +1$ and $f(x) < 0$ for $y_i = -1$. In SVM, the classifying function is

$$f(x) = \sum_{l=1}^{n} w_l \emptyset_l(\boldsymbol{x}) = w^T \emptyset(\boldsymbol{x}) + b$$

where $\emptyset_l(\boldsymbol{x}) = [\emptyset_1(\boldsymbol{x}), \dots, \emptyset_n(\boldsymbol{x})]^T$ is the set of nonlinear functions transforming the input space into a higher dimensional feature space. $w = [w_1, \dots, w_n]^T$ is the set of weights and b is the bias or threshold. The optimal w^* and b^* obtained from the training dataset is used to construct the optimal hyperplane $w^{*T}\emptyset(\boldsymbol{x}) + b^* = 0$ and classifying function $f(x) = w^{*T}\emptyset(\boldsymbol{x}) + b^*$. We can then use this $f(x)$ to perform prediction on the testing dataset.

In the linearly separable situation, there exists a perfect hyperplane in the input space and all data points satisfy $y_i \cdot f(x) \geq 1$. Data points for which the equality holds are called support vectors, located on the upper or lower separation band. Define the margin of separation to be the distance between the upper and lower separation band. The SVM algorithm simply aims to maximize the margin, which is formulated as $\max_{w,b} \frac{2}{\|w\|}$.

In the linearly non-separable situation, no perfect hyperplane can be found. An error term ξ is therefore introduced to relax the constraints, and the objective function is updated accordingly as

$$\min_{w,b,\xi} \frac{1}{2}\|w\| + C\sum_{i=1}^{n} \frac{\xi i}{\|w\|}$$

In the new objective function, a penalty parameter C is introduced, which controls the tradeoff between margin maximization and misclassification error minimization. The value of C is decided by the researcher (Haykin, 2001). The optimal $w^*$and b^* are derived from the solution to the corresponding dual problem since it is easier to solve than the primal (Tian & Deng, 2004). In the dual problem, an inner product kernel function $K(x_i, x_j) = \emptyset^T(x_i)\emptyset(x_j)$ is introduced, which is a crucial part of SVM because it effectively simplified the computation in high dimensions. Common kernels include linear, polynomial and Gaussian kernels. Our study employs the Gaussian kernel $K(X_i, X_j) = \exp(-(X_i - X_j)^T r^{-2} \sum(X_i - X_j)/2)$, where the Gaussian coefficient r controls algorithm complexity. Thus, two parameters, C and r, are to be determined.

SVM adopts the principle of structural risk minimization and also allows for incorporation of nonlinearity; therefore, it systematically outperforms other classification methods, particularly when the data is linearly non-separable or out-of-sample.

3. Literature Review

The application of SVM has been recently introduced to corporate bankruptcy analysis due to its superior classifying ability. Cizek, Härdle and Weron (2005) reported the superior performance of SVM in comparison with DA and logit model in corporate default prediction. Van Gestel et al. (2005) combined SVM and the logit model to cater for both interpretability and predictability of credit scoring. In the study of Chen, Härdle, and Moro (2011), the SVM technique was adopted to predict the default risk of German firms using their financial ratios. They selected eight most powerful predictors and reported that SVM outperforms the benchmark logit model. Their methodology and empirical results are presented below.

The paper used a dataset containing 20,000 solvent and 1,000 insolvent firms in Germany over the period spanning from 1996 to 2002. 28 financial ratios were selected as candidate predictors, covering various categories including profitability, leverage, liquidity and activity. To remove the effects of outliers, the ratios were scaled as follows: if xi < q0.05 (xi), then set xi = q0.05 (xi); if xi > q0.95 (xi), then set xi = q0.95 (xi), where i = 1, 2, …, 28 and qα (xi) is the α-quantile of xi.

Accuracy ratio (AR) is used to evaluate model performance. It is calculated as the ratio of the area between a model's cumulative accuracy profile (CAP) curve and the random CAP curve to the area between the perfect CAP curve and the random CAP curve (see Appendix Figure 1).

The authors constructed a training dataset containing 387 insolvencies and 3,534 solvencies and a testing dataset containing 396 insolvencies and 6,049 solvencies. To estimate the distribution of AR and effectively reduce over-fitting, the study employed bootstrap in both training and testing datasets: each subsample consists of all the insolvencies and a random sample with the same number of the solvencies; such a re-sampling procedure is repeated for 30 times.

The paper used forward stepwise selection due to its relatively lower computational cost. The first stage of model construction is to select the first best predictor using the median of the AR metric in which the SVM model has one input. The second stage is the selection of penalty parameter C and Gaussian kernel coefficient r to achieve an appropriate trade-off between type I and type II errors, which were calculated from single-predictors models. The final step is to add the financial ratio whose inclusion gives the highest AR, and repeat the process until none improves the model performance.

Finally, C = 10 and r = 0.6 were selected for default analysis. As shown in Table 1 in Appendix, the final model includes eight financial ratios indicating profitability, leverage, liquidity, activity, asset size and incremental inventories. The model gives an AR of 60.51, significantly outperforming the benchmark logit model with AR = 35.15.

4. Methodology

Based on the study of Chen, Härdle and Moro, we improved their prediction framework by taking the correlation among candidate predictors and class-depend misclassification costs into consideration.

4.1. Data Description

The dataset used in our study is from Center for Machine Learning and Intelligent Systems of University of California Irvine (UCI, n.d.). It contains financial ratios of 9,277 solvent and 515 insolvent companies in Poland spanning the period from 2003 to 2011. In accordance to the existing literature, 23 financial ratios are selected for default analysis, covering profitability, leverage, liquidity, activity, asset size and percentage change in inventories/liabilities/cash flow (see Appendix Table 2). To remove the effects of outliers, we scale the data in the same way as Chen, Härdle and Moro (2011).

4.2. Correlation of Predictors

We notice that some candidate predictors have similar financial meanings. For example, intuitively there is a strong positive correlation between liquidity indicators quick ratio (x15) and current ratio (x16), as the only difference between them is that quick ratio does not include inventory and other current assets that are relatively more difficult to liquidate (Folger, 2014). Adding highly correlated predictors to the model, however, would not improve the prediction accuracy. Worse still, it would increase computational cost. Even in the prediction model built by Chen, Härdle and Moro (2011), some predictors are likely to be highly correlated, such as net profit margin (x2) and EBIT/total asset (x7). Therefore, we consider dropping some candidate predictors before the model construction process.

4.3. Misclassification Cost

4.3.1. Objection

The objection of our classification model is to minimize the expected cost. Altman et al. (1997) estimated that the relative cost to commercial banks for misclassifying defaults loans into non-defaults (c1) and the opposite (c2) are calculated as follows:

$$c_1 = 1 - \frac{LLR}{GLL}, \quad c_2 = \text{r} - \text{i},$$

where:

LLR = loan losses recovered

GLL = gross loan losses

r = effective interest rate on the loan

i = effective opportunity cost for the bank

Empirical study shows that c1 is around 70% and c2 is around 2% (Altman et al.,1997). Therefore, the ratio of c1 and c2 is approximately 35:1. The objection function becomes

$$\min P(1|0) \cdot P(0) \cdot c_1 + P(0|1) \cdot P(1) \cdot c_2$$

where '1' denotes default loans and '0' denotes non-default loans.

4.3.2. Decision Threshold

SVM offers a binary output with a probability involved in the model. For binary problems with outputs 0 and 1, if the probability of being 1 is larger than the decision threshold, we accept it as 1. Otherwise, we accept it as 0. By default, p = 0.5 is used as the decision threshold. In order to incorporate different misclassification costs, the decision threshold could be adjusted.

4.3.3. Weighted Support Vector Machines

The main idea of weighted support vector machine (WSVM) is to assign different weights to different types of misclassifications such that the decision surface is learned according to the relative cost of different error types (Yang, Song & Wang, 2007). The objective function in the SVM training process therefore becomes:

$$\min_{w,b,\xi} \frac{1}{2} \parallel w \parallel^2 + \sum_{i=1}^{n} C_i \xi_i$$

s.t. $$y_i \times \{w^T \phi(x_i) + b\} + \xi_i \geq 1, \ \xi_i \geq 0, i = 1, 2, \ldots, n.$$

As shown above, C_i are the costs of different types of misclassification.

5. Results

5.1. Selection of the First Predictor

The first step of predictor selection is to exclude certain highly-correlated variables according to their financial meanings. As is shown in the correlation heatmap of the 23 candidate predictors (see Appendix Figure 2), all the seven predictors indicating profitability (x1 to x7) are highly correlated. We choose two commonly used factors, Net Profit Margin (x2) and EBITDA (x6), and drop the rest. Likewise, we exclude Own Funds Ratio (x8) and Current Liability/Total Assets (x10) in the leverage category and Cash Ratio (x14) and Current Ratio (x16) in the liquidity category. 14 variables remain as the inputs for SVM.

The performance of SVM with one input is illustrated in Table 1. The model is obtained using the training dataset. The fitted model obtained from the training dataset is then used to preform prediction on the testing dataset, during which AR is calculated.

Ratio No.	Ratio	AR Median
X2	Net profit margin	0.3982
X6	EBITDA	0.2770
X9	Sales(n)/Sales(n-1)	0.2379
X11	Net indebtedness	0.2868
X12	Total Liability/Total Assets	0.3446
X13	Cash/Total Assets	0.2082
X15	**Quick ratio**	**0.4141**
X17	Working Capital/Total Assets	0.3166
X18	Current Liability/Total Liability	-0.1204
X19	Asset turnover	-0.0032
X20	Inventory turnover	0.0867
X21	Account receivable turnover	-0.0216
X22	Account payable turnover	0.3475
X23	Log (Total Assets)	-0.0686

Table 1: Selection of the first predictor

Quick Ratio (x15) is the most significant predictor with which the model generates the highest median of AR. The result is reasonable in the sense that quick ratio is an important indicator of liquidity, which measures the firm's ability to pay off debts and is directly related to the probability of bankruptcy.

5.2. Optimal Values of C and r

As we have mentioned, there are two control parameters embedded in SVM which largely effect the model performance. Choosing proper values of the parameters will contribute to better model performance. However, as the optimal values of C and r could not be determined explicitly, they could only be derived by tuning. Here we use the most significant predictor selected in 5.1. as the model input and adjust the values of C and r to make the prediction error rate as small as possible. The sensitivity of error rate to different parameter values is illustrated in Appendix Figure 3. The optimal value of parameters is $C = 8$ and $r = 0.05$. The performance of the one-input model using the optimal C and r is shown below in Table 2. Compared with Table 1, the performance is better with general improvement in AR values. We will set the value of C and r to be the optimal values in the following studies.

Ratio No.	Ratio	AR Median
X2	Net profit margin	0.4318
X6	EBITDA	0.2735
X9	Sales(n)/Sales(n-1)	0.2378
X11	Net indebtedness	0.3093
X12	Total Liability/Total Assets	0.3661
X13	Cash/Total Assets	0.1933
X15	**Quick ratio**	**0.4373**
X17	Working Capital/Total Assets	0.3157
X18	Current Liability/Total Liability	-0.1207
X19	Asset turnover	-0.0589
X20	Inventory turnover	0.2219
X21	Account receivable turnover	-0.0588
X22	Account payable turnover	0.3556
X23	Log (Total Assets)	-0.0958

Table 2: Performance of one-input model with optimal *C* and *r*

5.3. Stepwise Parameter Selection

Forward stepwise selection was implemented for predictor selection using AR as criteria. First, we fix Quick Ratio (x15) and add another predictor to the input. The second predictor is selected by choosing the predictor whose inclusion produces the highest AR. The process was repeated until AR starts to decline. As shown in the line chart (see Appendix Figure 4), the best model includes six predictors with the highest AR equivalent to 0.6026. The six predictors are: Quick Ratio (x15), Net Profit Margin (x2), EBITDA (x6), Account Payable Turnover (x22), Log (Total Asset) (x23) and Inventory Turnover (x20).

5.4. Model Performance and Practical Concerns

Our final model includes six variables and has an AR equal to 0.6026. What is more, the error table (see Table 3) shows that 67% of the insolvent companies are predicted correctly.

Error Table

Prediction	Actual	
	Solvent	Insolvent
Solvent	66	33
Insolvent	34	67

Table 3

Recall that the existing methodology adopts a sampling method that enforces equal class sizes in both the training and testing dataset. However, if we test our SVM model in practical dataset where the number of insolvencies is much less than solvencies, no defaults are predicted correctly, as illustrated in the error table (see Table 4).

Error Table

Prediction	Actual	
	Solvent	Insolvent
Solvent	2,350	100
Insolvent	0	0

Table 4

To deal with the poor performance, we first try increasing the value of penalty parameter C to impose heavier penalty for misclassification. The error rate of the model in which the penalty is multiplied by 10,000 is shown in Table 5. The result is still unsatisfying in the sense that it only successfully predicts 4% of the insolvent cases at a cost of an increase in type I error (indicating low default risk when in fact the risk is high).

Error Table

Prediction	Actual	
	Solvent	Insolvent
Solvent	2,317	96
Insolvent	33	4

Table 5

Moreover, the computation gets much more complex and consumes much longer running time as C increases. Another problem of this method is that it imposes the same penalties for all misclassified points, regardless of their classes.

5.4.1. Decision Threshold

In our study, we adopt the conclusion of Altman et al. (1997) and set the cost of misclassifying insolvencies into solvencies to be 35 times of the opposite. Thus we decrease the decision threshold. If the default probability of a company is larger than a critical value p, which is less than 0.5, we will regard it as default. We train the model and find that the expected cost is minimized at $p = 0.037$. The error table is shown as follows:

Error Table

Prediction	Actual	
	Solvent	Insolvent
Solvent	50	0
Insolvent	2,300	100

Table 6

The probability heatmap for the first-two-predictor model is shown in Appendix Figure 5.

5.4.2. Weighted Support Vector Machines

As mentioned in 4.3.3., C1 and C2 are the penalties for two types of misclassification. We assign $C_i = c_1$ if misclassifying default loan into non-default, and $C_i = c_2$ in the other case. Let c1 be larger than c2, thus the penalty of fist type of misclassification is improved. We found when $c_1 : c_2 = 46:1$ the expected cost is minimized and that value is smaller than that by adjusting threshold. The error table is shown below:

Error Table

Prediction	Actual	
	Solvent	Insolvent
Solvent	1,313	19
Insolvent	1,037	81

Table 7

The probability heatmap for the first-two-predictor model is shown in Appendix Figure 6.

6. Conclusion

In this report, we have studied default risk modeling with SVM and built a prediction model for bankruptcy analysis of Polish firms. Our model includes six financial ratios with AR equivalent to 0.6026: quick ratio, net profit margin, EBITDA, account payable turnover, log (total asset) and inventory turnover. We have also incorporated class-dependent misclassification errors by adjusting the decision threshold and adopting the weighted SVM model. The weighted SVM model has better performance in minimizing the misclassification cost when the number of default and non-default data points is proportional to those in the real world.

In order to further analyze this problem, we may incorporate misclassification costs case by case. Specifically, different costs could be adjusted by the firm's recovery rate, asset size and loan amount or investors' risk appetite and investment horizon. We may also use other objective functions to train the model, such as to maximize the expected return.

References

Altman, E. I., Haldeman, R. G., & Narayanan, P. (1977). ZETATM analysis A new model to identify bankruptcy risk of corporations. *Journal of banking & finance*, *1*(1), 29-54.

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of accounting research*, *5*(suppl.): 71-111.

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

Cizek, P., Härdle, W. K., & Weron, R. (Eds.). (2005). “Predicting bankruptcy with support vector machines”. In *Statistical Tools for Finance and Insurance*. 225–248. Berlin: Springer science & business media.

Folger, J. (2014, September 02). *What are the main differences between the current ratio and the quick ratio?* Retrieved May 12, 2017, from http://www.investopedia.com/ask/answers/062714/what-are-main-differences-between-current-ratio-and-quick-ratio.asp

Haykin, S. S. (2001). *Neural networks: a comprehensive foundation*. Beijing: Tsinghua University Press.

Hertz, J., Krogh, A., & Palmer, R. G. (1991). *The Theory of Neural Network Computation.* Redwood CA: Addison Welsey.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research*, *18*: 109-131.

Tian, D. Y. Y. J., & Deng, N. (2004). *New methods of data mining: support vector machine.* Beijing: Science Press.

UCI. (n.d.). Retrieved May 12, 2017, from https://archive.ics.uci.edu/ml/datasets/Polish companies bankruptcy data

Van Gestel, T., Baesens, B., Van Dijcke, P., Suykens, J., Garcia, J., & Alderweireld, T. (2005). Linear and nonlinear credit scoring by combining logistic regression and support vector machines. *Journal of credit Risk*, *1*(4).

Vapnik, V. (1995). *The nature of statistical learning theory*. New York: Springer science & business media.

Yang, X., Song, Q., & Wang, Y. (2007). A weighted support vector machine for data classification. *International Journal of Pattern Recognition and Artificial Intelligence*, *21*(05), 961-976.

Appendix

Cumulative Accuracy Profile Curve

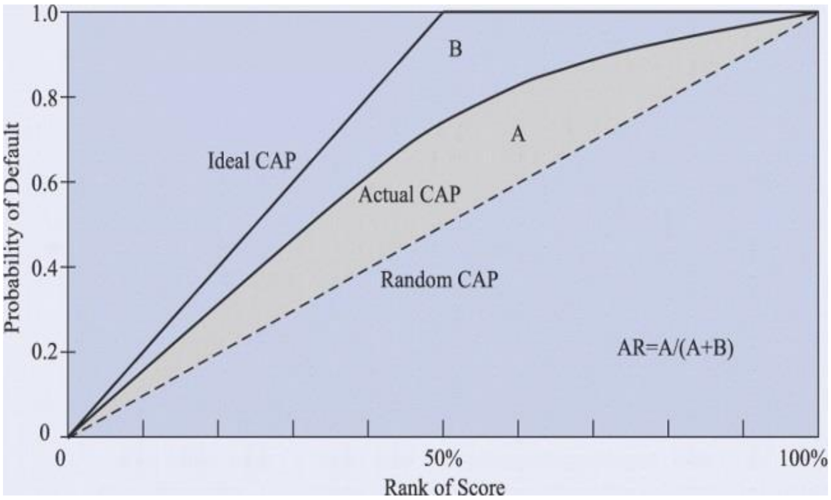

Figure 1: $AR = \frac{\int_0^1 y(x)dx - 1/2}{\int_0^1 y_{ideal}(x)dx - 1/2} = \frac{A}{A+B}$

No.	Ratio	Category
1	Account Payable Turnover	Activity
2	Operating Income/Total Assets	Profitability
3	Cash/Total Assets	Liquidity
4	Total Liabilities/Total Assets	Leverage
5	Incremental Inventories	Percentage
6	Inventory Turnover	Activity
7	EBIT/Sales	Profitability
8	Net Profit Margin	Profitability

Table 1: Predictors selected in the study of Chen, Härdle, and Moro (2011)

Category	No.	Ratio	Category	No.	Ratio
Profitability	x1	Return on assets (ROA)	Liquidity	x15	Cash/total asset
Profitability	x2	Net profit margin	Liquidity	x16	Cash ratio
Profitability	x3	Operating income/total asset	Liquidity	x17	Quick ratio
Profitability	x4	Operating profit margin	Liquidity	x18	Current ratio
Profitability	x5	EBIT/total asset	Liquidity	x19	Working capital/total asset
Profitability	x6	EBITDA	Liquidity	x20	Current liability/total liability
Profitability	x7	EBIT/sales	Activity	x21	Asset turnover
Leverage	x8	Own funds ratio (simple)	Activity	x22	Inventory turnover
Leverage	x9	Own funds ratio (adjusted)	Activity	x23	Account receivable turnover
Leverage	x10	Current Liability/total asset	Activity	x24	Account payable turnover
Leverage	x11	Net indebtedness	Size	x25	Log(Total asset)
Leverage	x12	Total liability/total asset	Percentage	x26	Percentage of incremental inventories
Leverage	x13	Debt ratio	Percentage	x27	Percentage of incremental liabilities
Leverage	x14	Interest coverage ratio	Percentage	x28	Percentage of incremental cash flow

Table 2: Candidate predictors in the study of Chen, Härdle, and Moro (2011)

Category	No.	Ratio	Category	No.	Ratio
Profitability	x1	Return on assets (ROA)	Liquidity	x14	Cash ratio
Profitability	x2	Net profit margin	Liquidity	x15	Quick ratio
Profitability	x3	Operating income/total assets	Liquidity	x16	Current ratio
Profitability	x4	Operating profit margin	Liquidity	x17	Working Capital/Total Assets
Profitability	x5	EBIT/Total Assets	Liquidity	x18	Current Liability/Total Liability
Profitability	x6	EBITDA	Activity	x19	Asset turnover
Profitability	x7	EBIT/Sales	Activity	x20	Inventory turnover
Leverage	x8	Own funds ratio (simple)	Activity	x21	Account receivable turnover
Leverage	x10	Current Liability/Total Assets	Activity	x22	Account payable turnover
Leverage	x11	Net indebtedness	Size	x23	Log(Total Assets)
Leverage	x12	Total Liability/Total Assets	Percentage	x9	Sales(n)/Sales(n-1)
Liquidity	x13	Cash/Total Assets	--	--	--

Table 3: Candidate predictors in our study

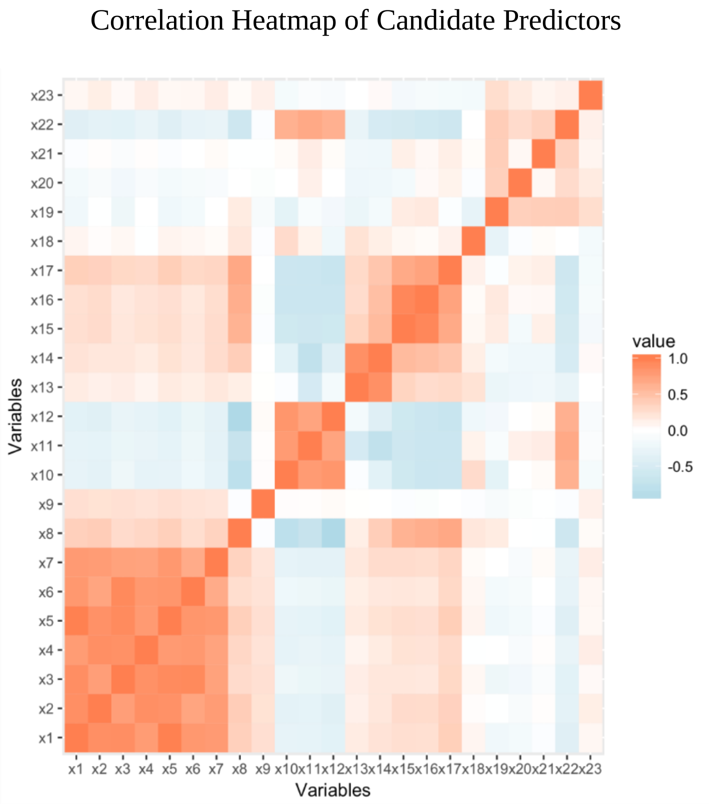

Figure 2

(A)

(B)

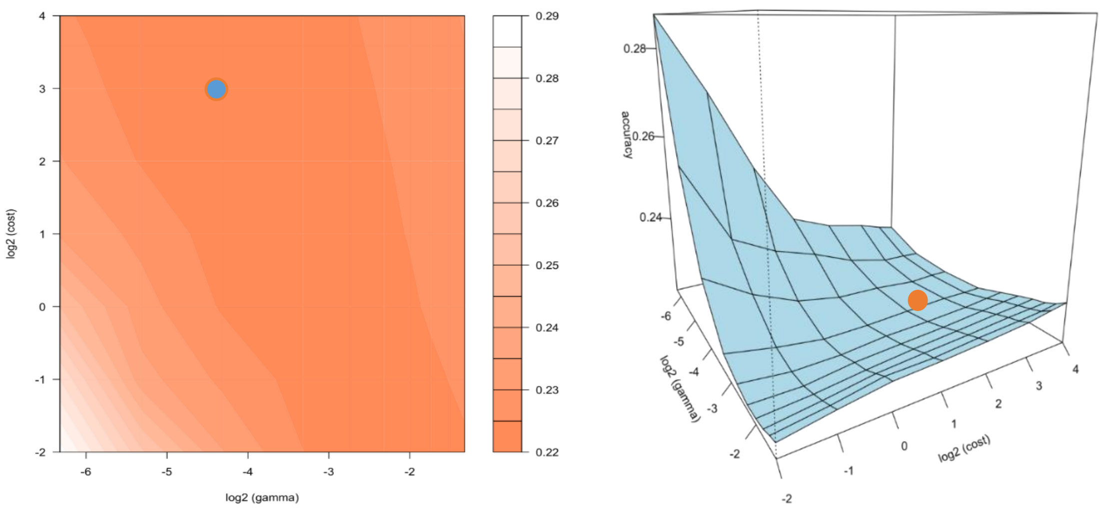

Figure 3: (A) The smallest error rate is represented by the darkest region. (B) The smallest error rate is represented by the lowest point.

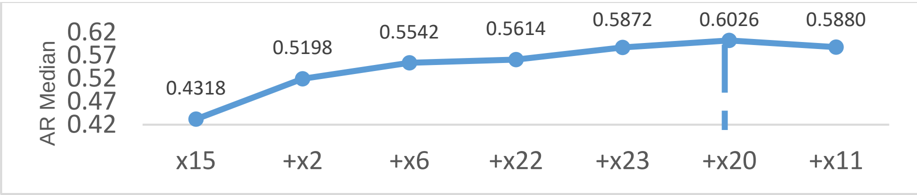

Figure 4: The highest AR is achieved when six predictors are included in SVM

Probability Heatmap for the Two-input Model with p = 0.037

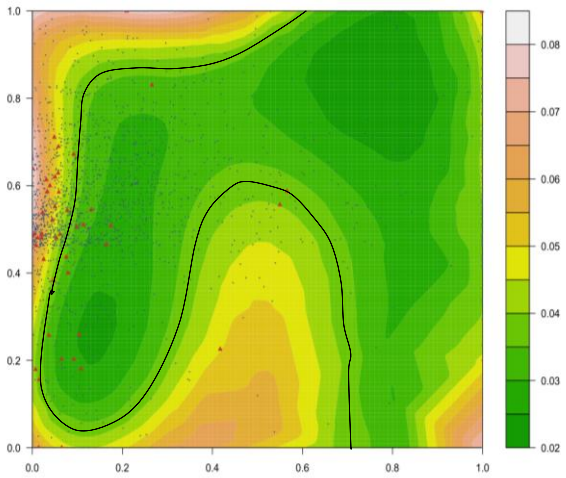

Figure 5: Red triangles indicate default firms and blue dots indicates non-default firms. The black line represents the decision boundary.

Probability Heatmap for the Two-input Weighted SVM Model

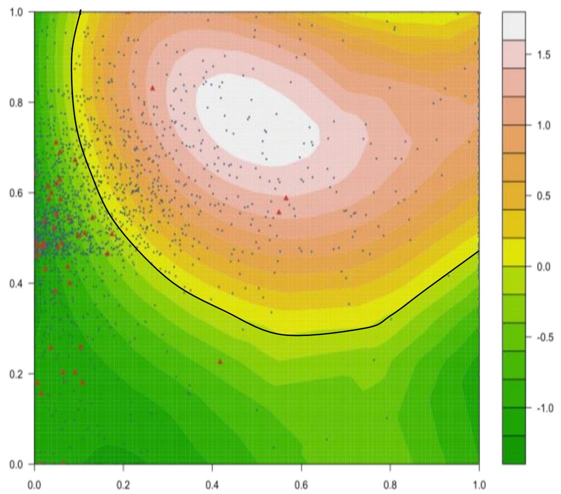

Figure 6: Red triangles indicate default firms and blue dots indicates non-default firms. The black line represents the decision boundary.