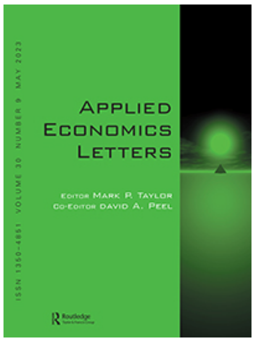

Routledge
Taylor & Francis Group

Applied Economics Letters

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rael20

Machine learning methods for predicting failures of US commercial bank

Le Quoc Tuan, Chih-Yung Lin & Huei-Wen Teng

APPLIED ECONOMICS LETTERS
https://doi.org/10.1080/13504851.2023.2186353

Routledge
Taylor & Francis Group

ARTICLE

Machine learning methods for predicting failures of US commercial bank

Le Quoc Tuan, Chih-Yung Lin and Huei-Wen Teng

Department of Information Management and Finance, National Yang Ming Chiao Tung University, Hsinchu, Taiwan

ABSTRACT
In this paper, we attempt to study the effectiveness of various simple machine learning methods in the prediction of bank failures. From a raw dataset of 10,938 US banks during the period of 2000–2020, we find that machine learning approaches do not really outperform the benchmark of conventional statistical method, logistic regression. However, using PCA to retain relevant variance in variables significantly improve the performance of machine learning methods and raise the out-of-sample accuracy of those method to over 70% to over 80%. Of all the machine learning methods used in this paper, the simple KNN seems to be the best model in forecasting bank failure in the United States.

KEYWORDS
Machine learning; bank failure; financial safety; banking system; financial indicators

JEL CLASSIFICATION
G21; C53

I. Introduction

Existing warning systems and models at time failed to predict the failure of financial institutions leading to the severe financial crisis of 2007–2009 (Davis and Karim 2008; Hasan, Liu, and Zhang 2016). As financial crisis is an economic-cycle matter, such crises will happen again. Fahlenbrach, Prilmeier, and Stulz (2012) show that banks that performed poorly in financial crisis of 1998 continue to do so in the 2008 crisis. Therefore, the development of better prediction method for failure of financial institutions is crucial to prevent or mitigate the damage caused by potential future crises.

While new assessment models have been developed following the 2007–2009 crisis (Oet et al. 2013; Shaddady and Moore 2019), recent studies started to use Information-Technology innovation for the job and find that more sophisticated machine learning methods could outperform conventional statistical models (Holopainen and Sarlin 2017; Samitas, Kampouris, and Kenourgios 2020).

While some models are being built for warning at country-level financial crisis; to the best of our knowledge, few studies have considered using machine learning at micro-economic level (e.g, bank level). Hence, we are interested in developing and comparing machine learning models attempting for better identification of bank failures based on selected financial indicators. We explore whether simple models could effectively identify troubled banks and what adjustments could be done to improve the efficiency of such models.

We collect the data of 10,938 US banks during a long period of 2000 to 2020 with around 30 financial indicators covering capital adequacy, asset and liabilities quality, profitability, and some off-balance sheet activities. We deploy logistic regression as a benchmark model for comparison of efficiency by simple machine learning models such as KNN, Decision Tree, and Random Forest. We find that using raw data of financial indicators, machine learning approaches do not really outperform the benchmark logistic model (except for slightly better in Random Forest).

Noticing that not all financial indicators are potentially useful in predicting bank failure (some create noises and disturbances in prediction), we deploy a Principal Component Analysis (PCA) to refine data by dimension reduction. The evidence shows that the adjustments significantly improve the performance of machine learning methods leading to a great outperformance to logit model.

We contribute to the growing literature of applying machine learning methods in modelling and

CONTACT Chih-Yung Lin d95723009@ntu.edu.tw Department of Information Management and Finance, National Yang Ming Chiao Tung University, Hsinchu, Taiwan

forecasting at micro-economic level (firms/banks, etc.). While proving such methods do not always outperform the conventional statistical approaches, we also show that their efficiency increases remarkably with appropriate adjustments.

II. Data and sample

We collect quarterly data on various financial indicators of US banks from the Bank Regulatory Database (Federal Reserve), available on Wharton Research Data Services. The database provides accounting data for bank holding companies, commercial banks, saving banks, and other S&L institutions. We define failure banks as either completely defunct or were acquired by other financial institutions during the sample period of around 20 years (2000–2020). To guarantee that the definition is proper, we also examine the list of failed banks provided by Federal Deposit Insurance Corporation.[1]

We follow previous studies (Shaddady and Moore 2019) and collect the important factors in predicting bank failure regarding capital adequacy, asset quality, liabilities quality, loan quality, earnings quality. We also consider off-balance-sheet activities as they are found by literature (Piskorski, Seru, and Vig 2010; Demyanyk and Hemert 2011) to be one of the main causes for the global financial crisis of 2007–2009. All variables and definitions are in Table 1.

Our final sample consists of 10,938 banks that ever existed during the sample period (approximately 400,000 observations), around 40% of which failed. Table 2 shows the summary statistics of all main variables.

III. Methodology

Our benchmark approach is logistic regression which is the conventional approach commonly used in predicting bank failure in previous literature (Martin 1977; Coles and Gunther 1995; Cleary and Hebbm 2016. The logistic model is as follows:

$Pr(Failure = 1|X) = \Lambda(X'\beta)$. In which, $\Lambda(X'\beta) = \frac{1}{1+exp^{-X'\beta}}$

X is the vector of independent variables mentioned in part 2.

The machine learning approaches that we considered in this paper are K-Nearest Neighbor (KNN), Decision Tree and Random Forest. A systematic overview on machine learning can be found in Hastie et al. (2009) and James et al. (2021).

K-Nearest Neighbour is a supervised learning algorithm where learning is based on how similar a data (vector) is from other. While it is considered a simple and 'lazy' machine learning approach, it is quite useful in binary classification and can estimate the nonlinear relationship while does not require assumptions about data. KNN is done by taking the *k* (odd number decided by user) nearest neighbours of the new data point to classified data (distance in Euclidian, Manhattan, Minkowski). Among these neighbours, the new data point is assigned to the category where the most neighbours are counted. In our study, Minkowski distance is chosen.

Decision Tree is a flowchart-like tree structure where an internal node represents feature/attribute (Breiman et al. 2017). A decision rule is represented by each branch and the outcome is depicted by each leafnote. The root node is the topmost node of a Decision Tree. The Decision Tree algorithm is set up by first selecting the best Attribute Selection Measures (ASM) and place at the root of the tree to start the splitting. Common ASMs are Information Gain, Gain ratio and Gini Index. Then, the training set is split into subsets. The process of tree building is repeated until one of the conditions matched: (i) all the tuples belong to the same attribute value, or (ii) more remaining attributes; or (iii) there are no more instances. In this study, Gini index is chosen:

$$Gini(D) = 1 - \sum_{i=1}^{m} p_i^2$$

Where pi is the probability that a tuple D belongs to class Ci. The attribute with minimum Gini index is chosen as the splitting attribute.

Random Forest consists of many individual Decision Trees that operate as an ensemble (Breiman 2001). Each individual tree results in a class prediction and those with the most votes

[1]https://www.fdic.gov/resources/resolutions/bank-failures/failed-bank-list/.

Table 1. Variables definition.

Variables	Definition
Outcome variable	
failure	Binary variable, equals one if the bank fails (defunct or acquired) and zero, otherwise
Features	
Bank assets	
size	The natural logarithm of bank's total assets (in millions of USD)
htminv_assets	Total investments in hold-to-maturity securities to total assets
fsale_assets	Total investments in securities available for sale to total assets
refarmcol_assets	Loan secured by farmland to total assets
interbank_assets	Loan to other banks to total assets
loanagri_assets	Loan to agriculture to total assets
clloan_assets	Commercial and Industrial loan to total assets
creditcard_assets	Credit card loan to total assets
otherresidloan_assets	Other loan to resident property to total assets
otherpploan_assets	Other personal loan to total assets
otherloan_assets	Total other loan to total assets
Bank liabilities	
lev	Total liabilities to total assets
demanddep_assets	Demand deposit to total assets
timedep_assets	Time deposit to total assets
Asset quality	
npl_assets	Delinquent loan to total assets
loanlossalwc	Loan loss allowance to total assets
loanlossprov	Loan loss provision to total assets
chargeoff_assets	Total loan charge-offs to total assets
Earnings quality	
nim	Net Interest Margin
nnon_im	Non-Interest Margin
roa	Returns on Assets
Capital adequacy	
tier1_car	Tier-1 ratio, calculated as tier-1 capital to Risk-weighted assets
Off-balance-sheet activities	
secborrowed_obs	Securities borrowed to total assets
seclent_obs	Securities lent to total assets
revolvelines_assets	Revolving Real Estate line of credit to total assets
creditcard_line_obs	Credit card line of credit to total assets
resid_recol_obs	Resident property line of credit secured by resident property to total assets
resid_norecol_obs	Resident property line of credit unsecured by resident property to total assets
underwriting_obs	Underwriting to total assets
derivatives_guarantor_obs	Derivatives with bank as guarantor to total assets
derivatives_beneficiary_obs	Derivatives with bank as beneficiary to total assets

become the model's prediction. The Random Forest algorithm: (i) random samples from a given dataset are selected; (ii), from each sample, a Decision Tree is constructed, and a prediction result is drawn; (iii) a vote for each predicted results is performed; and (iv), the prediction with the most votes is chosen as the final prediction. Random Forest is considered highly accurate and robust method based on wisdom of crowds from a committee of relatively uncorrelated models.

We split our full sample into training period and testing period for model construction and out-of-sample tests. The training period is the period of 2000–2010 while the testing period is defined as 2011–2020.

For assessment of different prediction methods and goodness-of-fit for each model, we deploy several metrics ratios from a confusion matrix, namely Accuracy; Precision; and Recall calculated from the number of correctly (incorrectly) identified failures, named True (False) Positive and the number of accurately (inaccurately) classified survival, namely True (False) Negative.

Table 2. Summary statistics.

Statistics	Obs.	Mean	Std	Minimum	p25%	p50%	p75%	Maximum
failure	379,783	0.4426	0.4967	0.0000	0.0000	0.0000	1.0000	1.0000
size	379,783	11.8810	1.6158	0.0000	10.8929	11.6722	12.5874	21.7132
htminv_assets	376,006	0.0394	0.0957	0.0000	0.0000	0.0000	0.0272	1.0000
fsale_assets	376,354	0.1805	0.1492	0.0000	0.0651	0.1558	0.2653	1.0000
lev	372,805	0.8876	0.0921	0.0000	0.8793	0.9035	0.9183	6.1947
refarmcol_assets	365,882	0.0366	0.0551	0.0000	0.0000	0.0102	0.0534	0.6494
interbank_assets	379,783	0.0002	0.0069	0.0000	0.0000	0.0000	0.0000	0.9146
loanagri_assets	357,345	0.0445	0.0796	0.0000	0.0000	0.0052	0.0529	0.7227
ciloan_assets	373,522	0.1014	0.0872	0.0000	0.0446	0.0810	0.1321	0.5066
creditcard_assets	379,783	0.0002	0.0008	0.0000	0.0000	0.0000	0.0000	0.0063
otherpploan_assets	353,545	0.0509	0.0495	0.0000	0.0154	0.0377	0.0700	0.2604
otherloan_assets	354,190	0.0025	0.0063	0.0000	0.0000	0.0004	0.0017	0.0427
demanddep_assets	375,431	0.1102	0.0766	0.0000	0.0619	0.1005	0.1443	1.0000
timedep_assets	354,190	0.6994	0.1234	0.0154	0.6623	0.7207	0.7696	0.8710
otherresidloan_assets	357,345	0.1488	0.1199	0.0000	0.0620	0.1235	0.2033	0.9825
npl_assets	371,643	0.0018	0.0044	0.0000	0.0000	0.0002	0.0017	0.2518
loanlossalwc	360,842	0.0092	0.0070	0.0000	0.0062	0.0082	0.0107	0.3673
loanlossprov	360,712	0.0020	0.0039	−0.0008	0.0001	0.0007	0.0020	0.0261
nim	354,626	0.0230	0.0115	0.0047	0.0120	0.0218	0.0315	0.0527
nnon_im	360,952	−0.0138	0.0095	−0.0477	−0.0190	−0.0126	−0.0070	0.0165
roa	361,026	0.0051	0.0084	−0.0360	0.0023	0.0051	0.0090	0.0302
tier1_car	355,693	0.1765	0.1563	0.0754	0.1088	0.1348	0.1832	1.3183
chargeoff_assets	357,344	0.0019	0.0123	0.0000	0.0001	0.0005	0.0016	5.9973
secborrowed_obs	357,330	0.0004	0.0094	0.0000	0.0000	0.0000	0.0000	0.7500
revolvelines_assets	357,345	0.0143	0.0241	0.0000	0.0000	0.0039	0.0199	0.9140
creditcard_line_obs	357,345	0.0049	0.0180	0.0000	0.0000	0.0000	0.0004	0.1575
resid_recol_obs	354,190	0.0267	0.0414	0.0000	0.0002	0.0123	0.0353	1.5345
resid_norecol_obs	357,345	0.0009	0.0061	0.0000	0.0000	0.0000	0.0000	0.5043
underwriting_obs	360,489	0.0000	0.0008	0.0000	0.0000	0.0000	0.0000	0.1210

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Accuracy shows the overall correctly identification of both positives and negatives. *Precision* is very important as it shows the number of correctly identified failure in all the predictions from our methods/approaches. If such number is low, that means our models provide many noisy predictions. In addition, if the regulators/policy-makers make decisions based on that, resources would be wasted on interventions/taking actions/restructuring the perfectly sound banks (FP). *Recall* shows that within all the truly failed banks, how many of them are correctly identified. This is also crucial considering that the bank failures are contagious and failing to identify them could proves fatal (Iyer and Peydro 2011).

IV. Results and analysis

Baseline results

In this section, we provide evidence of the prediction results for U.S bank failure by the benchmark and machine learning methods of choice. We first standardize all the main variables and run with raw data. The results are shown in Panel A of Table 3.

The evidence of logistics model shows an OK job in identifying bank failures. For both in-sample and out-of-sample test, the Accuracy metrics are over 60% which, unfortunately, mostly thanks to the accuracy of identifying survived banks. For in-sample test, the Precision of logistic regression is high (0.6261) while the Recall is lower (0.4625). In out-of-sample test, however, Precision of benchmark model greatly reduces to just over 20% while the Recall is nearly 55%.

For machine learning methods, the Accuracy, Precision and Recall metrics of in-sample test are extremely high which is normal since they often fit well with training data. However, overfitting could be a problem if those methods could not perform well for the testing data. Therefore, out-of-sample test comparison is more important in determining the goodness-of-fit of machine learning approaches. Interestingly, Panel A results indicate that simple machine learning models do not really outperform benchmark model as all goodness-of-fit metrics for KNN and Decision Tree are lower than logistic model. Only Random Forest method manages to get a higher Accuracy (70.76%) and Precision (23.89%) but lower Recall (0.407).

Table 3. Comparison of bank failure prediction methods: unadjusted data and after PCA.

Panel A: Unadjusted data				
In-sample test	Logistic regression	KNN	Decision Tree	Random Forest
Accuracy	0.6325	0.9459	1.0000	1.0000
Precision	0.6261	0.9422	1.0000	1.0000
Recall	0.4625	0.9378	1.0000	1.0000
Out-of-sample test				
Accuracy	0.6125	0.5689	0.5111	0.7076
Precision	0.2109	0.1847	0.1574	0.2389
Recall	0.5482	0.5227	0.4961	0.4070
Panel B: After PCA				
In-sample test				
Accuracy	0.6169	0.8578	1.0000	1.0000
Precision	0.6154	0.8544	1.0000	1.0000
Recall	0.4050	0.8260	1.0000	1.0000
Out-of-sample test				
Accuracy	0.5154	0.8109	0.7005	0.8057
Precision	0.1563	0.4372	0.2964	0.4278
Recall	0.4845	0.7735	0.6805	0.7558

The first evidence suggests that machine learning methods do not always outperform the traditional statistical model (like logit). This could also come from the fact that too many variables with potential redundant features could create noise that reduce the predicant power (Garbage in, Garbage out). Methods such as KNN are very sensitive to irrelevant features and missing data (Zhang 2020). Therefore, retaining only relevant features or variance of variables could potentially improve the prediction power of such machine learning methods.

PCA and results with robust data

PCA captures the intrinsic variability in the data and could be used to extract most representative features. Therefore, it is appropriate in this case where we are dealing with many features and some of them might be irrelevant. We deploy Scree plot and Eigenvalues to identify the proper number of dimensions to be used. The Scree plot is demonstrated in Figure 1.

The Scree plot shows that from the principal component 11, the proportion of variance explained stop changing significantly. The Eigenvalues of all features also confirm the use of 11 principal components whose eigenvalues greater than 1. Hence, we transform the data and run all the methods again using the new dataset and report the results in Panel B of Table 3.

The performance of benchmark logistic model did not change much with the new dataset and still provide above 50% accuracy in identifying failure banks. Meanwhile, using the new dataset proves to be extremely beneficial for the machine learning approach as all KNN, Decision Tree, and Random Forest method earn remarkable improved performance in bank failure prediction. For out-of-sample test, both Random Forest and KNN have over 80% Accuracy and over 75% Recall. The Precision of those methods also improves to over 40% and significantly higher than that of the benchmark method. Of three machine learning approaches, KNN seems to be the best model in terms of goodness-of-fit metrics in predicting bank failure.

V. Conclusions

In this study, we deploy various machine learning methods; namely, K-Nearest-Neighbour, Decision Tree, and Random Forest; to predict the failures of U.S banks during the period of more than 20 years (2000 to 2020). With the raw data of around 30 features from different aspects of bank operation, the machine learning approaches do not seem to produce better failure prediction of banks than the benchmark approach in out-of-sample tests. However, using PCA to transform the data and retain relevant principal components significantly improve the accuracy and prediction power of machine learning approaches to the point that

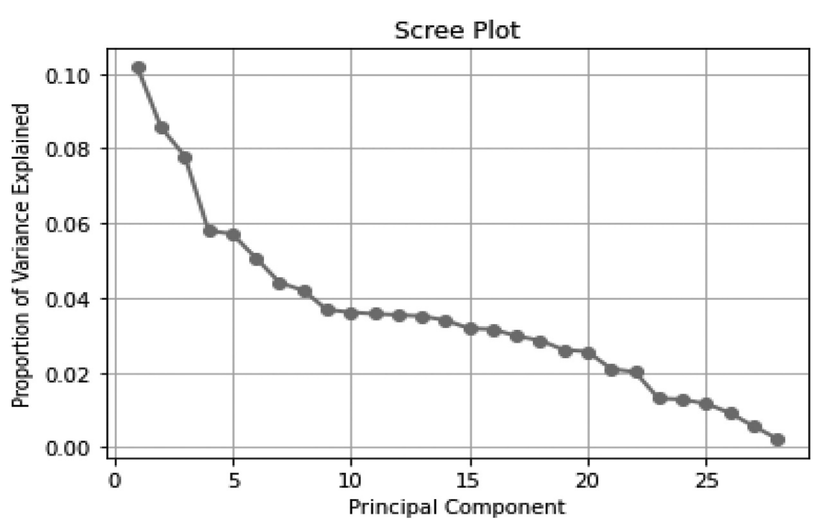

Figure 1. Scree plot for PCA.

they greatly outperform logistic regression for both in-sample and out-of-sample data.

Despite being a simple approach, the efficiency of KNN is surprisingly high. Whitin the other two approaches in this assignment, Random Forest shows its superiority of wisdom of crowds in comparison to Decision Tree with overwhelming accuracy and precision of predicting failure among US banks.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data subject to third party restrictions

The data that support the findings of this study are available from Compustat and CRSP. Restrictions apply to the availability of these data, which were used under licence for this study

References

Breiman, L. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. doi:10.1023/A:1010933404324.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 2017. *Classification and Regression Trees*. New York: Routledge.

Cleary, S., and G. Hebbm. 2016. "An Efficient and Functional Model for Predicting Bank Distress: In and Out of Sample Evidence." *Journal of Banking & Finance* 64: 101–111. doi:10.1016/j.jbankfin.2015.12.001.

Coles, R. A., and J. W. Gunther. 1995. "Separating the Likelihood and Timing of Bank Failure." *Journal of Banking & Finance* 19 (6): 1073–1089. doi:10.1016/0378-4266(95)98952-M.

Davis, E. P., and D. Karim. 2008. "Comparing Early Warning Systems for Banking Crises." *Journal of Financial Stability* 4 (2): 89–120. doi:10.1016/j.jfs.2007.12.004.

Demyanyk, Y., and O. V. Hemert. 2011. "Understanding the Subprime Mortgage Crisis." *The Review of Financial Studies* 24 (6): 1848–1880. doi:10.1093/rfs/hhp033.

Fahlenbrach, R., R. Prilmeier, and R. M. Stulz. 2012. "This Time is the Same: Using Bank Performance in 1998 to Explain Bank Performance During the Recent Financial Crisis." *The Journal of Finance* 67 (6): 2139–2185. doi:10.1111/j.1540-6261.2012.01783.x.

Hasan, I., L. Liu, and G. Zhang. 2016. "The Determinants of Global Bank Credit-Default-Swap Spreads." *Journal of Financial Services Research* 50 (3): 275–309. doi:10.1007/s10693-015-0232-z.

Hastie, T., R. Tibshirani, J. H. Friedman, and J. H. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer.

Holopainen, M., and P. Sarlin. 2017. "Toward Robust Early-Warning Models: A Horse Race, Ensembles and Model Uncertainty." *Quantitative Finance* 17 (12): 1933–1963. doi:10.1080/14697688.2017.1357972.

Iyer, R., and J. L. Peydro. 2011. "Interbank Contagion at Work: Evidence from a Natural Experiment." *The Review of Financial Studies* 24 (4): 1337–1377. doi:10.1093/rfs/hhp105.

James, G., D. Witten, T. Hastie, and R. Tibshirani. 2021. *An Introduction to Statistical Learning: With Applications in R*. 2 ed. New York, NY: Springer.

Martin, D. 1977. "Early Warning of Bank Failure, a Logit Regression Approach." *Journal of Banking & Finance* 1 (3): 249–276. doi:10.1016/0378-4266(77)90022-X.

Oet, M. V., T. Bianco, D. Gramlich, and S. J. Ong. 2013. "SAFE: An Early Warning System for Systemic Banking Risk." *Journal of Banking & Finance* 37 (11): 4510–4533. doi:10.1016/j.jbankfin.2013.02.016.

Piskorski, T., A. Seru, and V. Vig. 2010. "Securitization and Distressed Loan Renegotiation: Evidence from the Subprime Mortgage Crisis." *Journal of Financial Economics* 97 (3): 369–397. doi:10.1016/j.jfineco.2010.04.003.

Samitas, A., E. Kampouris, and D. Kenourgios. 2020. "Machine Learning as an Early Warning System to Predict Financial Crisis." *International Review of Financial Analysis* 71: 101507. doi:10.1016/j.irfa.2020.101507.

Shaddady, A., and T. Moore. 2019. "Investigation of the Effects of Financial Regulation and Supervision on Bank Stability: The Application of CAMELS-DEA to Quantile Regressions." *Journal of International Financial Markets, Institutions and Money* 58: 96–116. doi:10.1016/j.intfin.2018.09.006.

Zhang, S. 2020. "Cost-Sensitive KNN Classification." *Neurocomputing* 391 (28): 234–242. doi:10.1016/j.neucom.2018.11.101.