Corporate Risk Estimation by Combining Machine Learning Technique and Risk Measure

[1]Hsu, Yushan, [2]Hsu, Ming-Fu, [3*]Lin, Sin-Jin
[1]Administrator, Department of Surgery, Kaohsiung Municipal Ta-Tung Hospital, Taiwan, R.O.C., yushanyushanh@gmail.com
[2]Department of Information Management, National Sun Yat-sen University, Taiwan, R.O.C., hsumf0222@gmail.com
[3]Department of Accounting, Chinese Culture University, Taiwan, R.O.C., annman1204@gmail.com (*Corresponding author)

***Abstract*—A precise measure of corporates' operating performance play critical role in it achieving sustainable development during turbulent financial markets, because operating performance is a suitable reflection of corporate management, which has been widely recognized as the main cause of financial troubles. However, most proportion of previous works take return on assets or return on investment as a proxy for operating performance assessment, but both assessing measures merely consider one input and one output features, making them not appropriate for describing the whole facets of a corporate's operating situation. These limitations suggest that there is a need for additional analytical model for effective and useful prediction of corporate operating performance ranking. Thus, this study proposes a reliable and sophisticated prediction architecture that incorporates risk metrics, dimensionality reduction technique, data envelopment analysis, and artificial intelligence technique for corporate operating performance forecasting. The experimental results show that the proposed architecture can reduce unnecessary information, satisfactorily forecast the corporate operating performance ranking, and yield directions for properly allocating limited financial resource on reliable objects. The introduced architecture is a promising alternative for predicting corporate operating performance ranking, it can assist in both internal and external decision makers.**

Keywords—risk management; decision making; artificial intelligence; performance measure

I. Introduction

The global economic crisis that erupted in 2007 turned the spotlight onto corporates and their risk exposures. A fundamental and timely question is how the risk of a corporate should be measured [1-2]. This study proposes an emerging architecture to measure corporate's risk utilizing the variance of the performance function. We model risk as the variance of the performance function, where the variance enters as a multiplicative component of the error term [3]. Grounded on the theory of accounting and corporate finance, the corporate make risky decisions simultaneously with the perception about the expected profits and of the level of other corporate characteristics, mainly debt and capital structure [4-5].

Financial ratios are a conventional reliable way to evaluate corporate operating performance. Rather than implementing all the information observed in financial documents, these measures are examined to obtain meaningful results [6]. The results provide appropriate measurements for high-level managers or investors to assess their firm's operating situation. However, most proportion of previous works take return on investment (ROI) or return on equity (ROE) as a proxy for operating performance evaluation, but both indicators merely consider one input and one output variable, making them not appropriate for describing the whole facets of a corporate's operating situation. Thus, the reliable and sophisticated techniques are required urgently.

Data envelopment analysis (DEA), proposed by Charnes et al. [7], is a non-parametric approach for evaluating the relative efficiency of a set of similar units, usually referred to as decision-making units (DMUs). It converts multiple input and multiple output variables [8-9] and assumes neither a specific functional form for the production function nor an inefficiency distribution, in comparison with the parametric statistical approach. Due to its superior merits, it has been extended to numerous research problems. However, one of the critical challenges of DEA is lacking of discriminant ability when there is a relatively large amount of variables as compared to DMUs [10]. To tackle with the problem encountered by DEA, the dimensionality reduction approach, namely random projection (RP), is utilized. It projects a set of points from a high-dimensional space onto a randomly low-dimensional subspace and does not significantly distort data [11-12]. Furthermore, it has also been demonstrated as being computationally efficient and a sufficiently precise technique for the task of dimensionality reduction. The information retrieval and dimensionality reduction power of RP is solidly proven by the theory of compressed sensing (CS) [13-14], which indicates that for sparse and compressible signals, a small number of non-adaptive linear measurements in the form of RPs can acquire the most representative information in the signal [14]. Thus, RP can handle the curse of dimensionality as well as improve the discriminant ability of DEA.

One of the emerging machine learning techniques, namely extreme learning machine (ELM), has been introduced recently [15]. In comparison with traditional neural network-based techniques, such as back propagation neural network (BPNN) and support vector machine (SVM), ELM has a much faster training speed and preferable generalization capability. Furthermore, ELM solves numerous problems encountered by gradient-based learning algorithm, such as learning rate, stopping criteria, number of epochs, and local minima [16]. Due to the aforementioned advantages, ELM was taken as a prediction architecture in this study.

The rest of this study is structured as follows. Section 2 expresses research methodologies. Section 3 presents the empirical outcomes. Section 4 concludes and provides future research directions.

ICIS 2016, June 26-29, 2016, Okayama, Japan

II. METHODOLOGIES

A. Data envelopment analysis and Random projection

Data envelopment analysis (DEA) is a systematic method for measuring the efficiency of a decision-making unit (DMU). To decide the relative efficiency of each DMU in a homogeneous category, DEA collapses the input variables and output variables determined by the mechanism into a ratio of a single meta-input and meta-output. It then implements mathematical programming approaches to provide the efficiency score for each DMU, where the provided score is used to represent the relative performance [17-18].

Adler and Yazhemsky [19] stated that a problem related to discrimination arises - for example, when there is a relatively large amount of variables as compared to DMUs, which in extreme cases may cause the majority of observations to be deemed as efficient. Thus, they performed a dimensionality reduction technique to solve the related problems as well as to improve DEA's discriminant ability. Random projection, one type of dimensionality reduction method, is a linear technique for the projection from a high-dimensional space into a low-dimensional space, utilizing projection vectors with random numbers as vector components [20]. The basic idea of RP is that pairs of high-dimensional vectors with randomly decided vector components have a high possibility of being "almost orthogonal" [21]. RP transforms a statistical unbiased sampling of the high-dimensional space into a tractable low-dimensional one, where properties like distances between points and certain other features are conserved as defined by the Johnson-Lindenstrauss lemma [22-23]. As RP is simple and efficient in computation, this study adopts it.

B. Extreme learning machine

Extreme learning machine (ELM) was developed for SLFNs and is admired for its superior training speed, because it adopts random hidden node parameters and calculates the output weights with the least square method [24-26]. With these advantages, ELM has overcome numerous limitations existing in the traditional neural network model with a gradient-descent based algorithm. However, the amounts of hidden node and hidden node parameters embedded into ELM are still determined manually, which leads to the existence of non-optimal parameters and deteriorates the goal of minimizing the cost function [26]. Optimizing the neural network parameter can be handled by the differential evolution (DE) algorithm, which is one kind of evolutionary algorithm, is easy to implement, is an effective population-based stochastic direct searching method, and has been widely used in determining the network parameters. Thus, the DE was used in this study.

III. EMPIRICAL OUTCOMES

A. Data

The firms categorized into the electronic industries have been selected as the research samples, because the amount of all stock transactions in those specific industries is on average over 70% of Taiwan's daily stock market turnover. These specific industries are the mainstream and have considerable impact on the Taiwan stock market. All research data were collected from public websites: Taiwan Economic Journal (TEJ) data bank, Taiwan Stock Exchange Corporation (TSEC), and Gre-Tai Securities Market (GTSM). The research period ranges from 2010 to 2015.

B. Features

In order to construct a prediction architecture for corporate operating assessment - which is ordinarily based on measuring the ratios gathered from financial documents such as income statement, balance sheet, etc. - the predictor variables should be decided at the beginning. The predictor variables are separated into two parts: decision variable and condition variable. The former is determined by DEA, and the latter is derived from financial reports. The feature for DEA was represented in Table 1. Corporate operating assessment is considerably related to the field of financial crisis prediction, whereby the selected features are taken as the candidates for joining the condition features.

TABLE I. THE FEATURES FOR DEA

Features for DEA	
Input feature	Output feature
I1:Total debts	O1: Earnings before interest and tax
I2: Total assets	O2: Sales revenue

TABLE II. THE DECISION FEATURES.

Features	**Description**
F1: TD/TA	Total debts to Total assets
F2: NI/TA	Net income to Total assets
F3: CA/TS	Current assets to Total sales
F4: CA/CL	Current assets to Current liabilities
F5: OI/TA	Operating income to Total assets
F6: QA/CL	Quick assets to Current liabilities
F7: CF/TD	Cash flow to Total debts
F8: I/TS	Inventory to Total sales
F9: LTD/TA	Long-term debts to Total assets

C. Assessing criteria

We investigate the usefulness of the proposed prediction architecture by utilizing a real dataset in Taiwan. The prediction error rate and accuracy are at the core for measuring the criteria in the prediction task. This study focuses on overall accuracy and two evaluating criteria, sensitivity and specificity, that take misclassification into consideration. Sensitivity is defined as the proportion of positive types that are forecasted to be positive; specificity is the proportion of negative types that are forecasted to be negative. All of the measuring criteria have been computed according to the confusion matrix (see Table 3), and the mathematical formats are described as follows.

Overall accuracy: (TN+TP)/(TP+FP+FN+TN)

Sensitivity: (TP)/(TP+FN)

Specificity: (TN)/(FP+TN)

TABLE III. THE CONFUSION MATRIX

Situation	**Predicted positive**	**Predicted negative**
Positive case	True positive (TP)	False negative (FN)
Negative case	False positive (FP)	True negative (TN)

Identify applicable sponsor/s here. If no sponsors, delete this text box (*sponsors*).

D. Research outcome

To alleviate the problem of over-fitting, this study performs k-fold cross-validation, in which we initially divide the training dataset into subsets of equivalent size. One subset is then investigated using the model trained on the remaining (k-1) subsets. Under this process, each instance of the whole training set is forecasted once, and so the cross-validation performance is the percentage of data that are correctly classified. To examine the suitability of the selected features for DEA, we implement the Pearson correlation that demonstrates significant positive correlations among all selected features (see Table 4). Thus, the selected features undergoing the random projection technique are fed into DEA to rank corporate operating efficiency. To take the risk factor into consideration, we utilize the variation in performance score as a more comprehensive risk measurement [3]. That is, we use information from a fixed number of periods to calculate the variance of performance score or the coefficient of variation as a measure of risk (DEA/σ_{DEA}) [27-28]. The final outcome is the risk-adjusted performance situation. The performance scores obtained from risk-adjusted performance assessment in the highest ranking quintile (top 20%) denote efficient corporates, while those in the lower four quintiles (bottom 80%) of performance ranking are taken as inefficient corporates.

TABLE IV. THE RESULT OF PEARSON CORRELATION

	I1: TL	I2: TA	I3: CoS	O1: NI	O2: TS
I1: TL	1				
I2: TA	0.985	1			
I3: CoS	0.913	0.913	1		
O1: NI	0.687	0.687	0.687	1	
O2: TS	0.714	0.714	0.714	0.714	1

Note: all coefficients are statistically significant at the 1% level.

To make the prediction performance more robust, this study takes the introduced architecture as a benchmark and compares it with other classifiers - namely, support vector machine (SVM), relevance vector machine (RVM), back-propagation neural network (BPNN), and decision tree (DT) - under two dissimilar scenarios (Scenario 1: With dimensionality reduction; Scenario 2: Without dimensionality reduction) (see Tables 5-6).

TABLE V. THE FORECASTING RESULTS

Multiple models comparison under Scenario 1: with dimensionality reduction			
Model	***Assessing criteria (Avg.)***		
	Overall accuracy	***Sensitivity***	***specificity***
Proposed architecture	93.04	87.45	94.44
SVM	89.89	80.15	92.32
RVM	88.04	77.55	90.66
BPNN	88.14	75.40	91.33
DT	85.62	71.80	89.08
Average performance score: 0.643			

TABLE VI. THE FORECASTING RESULTS

Multiple models comparison under Scenario 2: without dimensionality reduction			
Model	***Assessing criteria (Avg.)***		
	Overall accuracy	***Sensitivity***	***Specificity***
Proposed architecture	87.59	76.95	90.25
SVM	83.94	68.40	87.83
RVM	81.19	64.55	85.85
BPNN	81.59	61.15	86.20
DT	78.17	56.30	83.64
Average performance score: 0.784			

Furthermore, the majority of prior studies merely utilize one pre-decided dataset to make a final judgment. To avoid making an inappropriate decision and unreliable judgment, this study considers another financial crisis dataset provided by Pietruszkiewicz [29] for evaluation. Table 7 shows the results. The proposed architecture still outperforms the other models.

TABLE VII. THE FORECASTING RESULTS IN POLAND FINANCIAL CRISIS DATABASE.

Model	**Assessing criteria (Avg.)**		
	Overall accuracy	***Sensitivity***	***specificity***
Proposed architecture	90.17	93.75	87.03
SVM	82.25	87.14	77.97
RVM	78.75	80.54	77.19
BPNN	77.42	80.18	75.00
DT	74.42	79.29	70.16

IV. CONCLUSIONS

This study proposes the merging risk-adjusted performance prediction model for decision makers to achieve the performance level in a homogeneous industry and to generate reliable operating strategies so as to achieve the goal of sustainable development in a highly fluctuated environment. Corporate operating performance is a good reflection of a corporate's management, which has been widely recognized to be the main reason of financial troubles. Thus, DEA is performed as a tool to evaluate the multiple input/output efficiency of each corporate. However, too many features for DEA will deteriorate its discriminant ability. Random projection technique, one kind of dimensionality reduction technique, was implemented to deal with the aforementioned problem. Furthermore, this study advanced take risk exposure into consideration in order to reach a more comprehensive risk measurement.

Inspired by hybrid mechanism, the introduced architecture herein, based on DE-ELM, predict the corporate operating situation. The hybrid mechanism has been proven its outstanding prediction performance. From the top-managers' point of view, the introduced model can be taken as guideline to modify the corporate operating strategy and adjust future directions. From the viewpoint of investors, the model can be used as pre-warning system to protect their personnel wealth as well as sound the stability of financial market.

ACKNOWLEDGMENT

The author would like to thanks Ministry of Science and Technology (MOST) of the Republic of China, Taiwan for financially supporting this work under Contract No.104-2410-H-034 -023 -MY2.

REFERENCES

[1] Lee, M., and Wu, T.C., 2012, "A integral predictive model of financial distress," J. Test. Eval. Vol. 40, pp. 931-938.

[2] Lee, L., Fan, C., Hung, H., and Ling, Y., 2010, "Analysis of financial distress prediction model," J. Test. Eval. Vol. 38, pp. 840-847.

[3] Delis, M.D., Hasan, I., and Tsionas, E.G., 2014, "The risk of financial intermediaries," J. Bank. Financ. Vol. 44, pp. 1-12.

[4] Kim, D., and Santomero, A.M., 1988, "Risk in banking and capital regulation," J. Financ. Vol. 43, pp. 1219-1233.

[5] Hughes, J.P., Mester, L.J., and Moon, C.G., 2001, "Are scale economies in banking elusive or illusive? Evidence obtained by incorporating capital structure and risk-taking into models of bank production," J. Bank. Financ. Vol. 25, pp. 2169–2208.

[6] Delen, D., Kuzey, C., and Uyar, A., 2013, "Measuring firm performance using financial ratios: A decision tree approach," Expert Syst. Appl. Vol. 40, pp. 3970-3983.

[7] Charnes, A., Cooper, W.W., and Rhodes, E., 1978, "Measuring the efficiency of decision making units," Eur. J. Oper. Res. Vol. 2, pp. 429-444.

[8] Luo, Y., Bi, G., and Liang, L., 2012, "Input/output indicator selection for DEA efficiency evaluation: An empirical study of Chinese commercial banks," Expert Syst. Appl. Vol. 39, pp. 1118-1123.

[9] Wang, W.K., Lu, W.M., Kweh, Q. L., and Cheng, I.T., 2014, "Does intellectual capital matter? Assessing the performance of CPA firms based on additive efficiency decomposition DEA," Knowl.-Based Syst. Vol. 65, pp. 38-49.

[10] Adler, N., and Golany, B., 2001, "Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to Western Europe," Eur. J. Oper. Res. Vol. 132, pp. 260-273.

[11] Amador, J.J., 2007, "Random projection and orthonormality for lossy image compression," *Image Vision Comput.* Vol. 25, pp. 754-766.

[12] Liu, L., Fieguth, P., Clausi, D., and Kuang, G., 2012, "Sorted random projections for robust rotation-invariant texture classification," *Pattern Recogn.* Vol. 45, pp. 2405-2418.

[13] Candès, E.J., and Tao, T., 2006, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE T. Inform. Theory*. Vol. 52, pp. 5406-5425.

[14] Tsaig, Y., and Donoho, D.L., 2006, "Extensions of compressed sensing," *Signal Process.* Vol. 86, pp. 549-571.

[15] Huang, G.B., Zhu, Q.Y., ans Siew, C.K., 2006, "Extreme learning machine: Theory and applications," *Neurocomputing*. Vol. 70, pp. 489-501.

[16] Li, K., Kong, X., Lu, Z., Wenyin, L., and Yin, J., 2014, "Boosting weighted ELM for imbalanced learning," *Neurocomputing*, Vol. 128, pp. 15-21.

[17] Shimshak, D.G., Lenard, M.L., and Klimberg, R.K., 2009, "Incorporating quality into data envelopment analysis of nursing home performance: A case study," *Omega*, Vol. 37, pp. 627-685.

[18] Bougnol, M.L., Dulá, J.H., Estellita Lins, M.P., and Moreira da Silva, A.C., 2010, "Enhancing standard performance practices with DEA," *Omega*, Vol. 38, pp. 33-45.

[19] Adler, N., and Yazhemsky, E., 2010, "Improving discrimination in data envelopment analysis: PCA–DEA or variable reduction," *Eur. J. Oper. Res.* Vol. 202, pp. 273-284.

[20] Vempala, S., 2004, "The random projection method," American Mathematical Society, Providence, RI.

[21] Kaski, S., 1998, "Dimensionality reduction by random mapping: fast similarity computation for clustering," Proceedings of the 1998 IEEE International Joint Conference on Neural Networks 413-418.

[22] Johnson, W.B., and Lindenstrauss, J., 1984, "Extensions of Lipschitz maps into a Hilbert space," *Contemp. Math*. Vol. 26, pp. 189-206.

[23] Varmuza, K., Engrand, C., Filzmoser, P., Hilchenbach, M., Kissel, J.H., Silén, J., and Trieloff, M., 2011, "Random projection for dimensionality reduction-Applied to time-of-flight secondary ion mass spectrometry data," *Analytica Chimica Acta*, Vol. 705, pp. 48-55.

[24] Huang, G.B., Chen, L., and Siew, C.K., 2006, "Universal approximation using incremental constructive feedforward networks with random hidden nodes," *IEEE T. Neural Networks*, Vol. 17, pp. 879-892.

[25] Lan, Y., Soh, Y.C., and Huang, G.B., 2009, "Ensemble of online sequential extreme learning machine," *Neurocomputing,* Vol. 72, pp. 3391-3395.

[26] Cao, J., Lin, Z., and Huang, G.B., 2012, "Self-adaptive evolutionary extreme learning machine," *Neural Process. Lett.* Vol. 36, pp. 285-305.

[27] DeYoung, R., and Rice, T., 2004, "Noninterest income and financial performance at U.S. commercial banks," *Financ. Rev.* Vol. 39, pp. 101–127.

[28] Stiroh, K.J., 2004, "Diversification in banking: is noninterest income the answer?," *Journal of Money, Credit and Banking,* Vol. 36, pp. 853–882.

[29] Pietruszkiewicz, W., 2008, "Dynamical systems and nonlinear Kalman filtering applied in classification," Proceedings of 7th IEEE International Conference on Cybernetic Intelligent Systems 263-268.