EVALUATING THE EFFICIENCY OF THE IT VENTURE BUSINESS USING DEA AND DATA MINING

Research-in-Progress

Taeho Hong
School of Business,
Pusan National University
30 Jangjeon-dong, Geumjeong-gu, Busan
609-735, Korea
hongth@pnu.edu

Jiyong Park
School of Business,
Pusan National University
30 Jangjeon-dong, Geumjeong-gu, Busan
609-735, Korea
jiyong@pnu.edu

Taewon Lee
School of Business, Pusan National University
30 Jangjeon-dong, Geumjeong-gu, Busan 609-735, Korea
twanny@pnu.edu

Abstract

Proposed in this paper, is a DEA-Based data mining approach. To predict the efficiency of an IT venture business, data mining was utilized, to discover advantageous patterns in data. To measure the efficiency of an IT venture business, the Data Envelopment Analysis (DEA) was employed. This is a typical approach among non-parametric methods for measuring the efficiency of companies. A selection of KOSDAQ firms was divided into two groups in accordance with the efficiency in the DEA model. Using a logit model through the stepwise, we finally acquired a model for evaluating the efficiency of an IT venture business. We applied our integrated model to companies listed on the KOSDAQ, which is a stock market division of Korea exchange for dealing the securities of the venture business, with the corporate information available from 2005. Our integrated model enabled us to evaluate an individual firm and provided efficiency information without comparing with other companies. In this paper, to examine the feasibility of SVM in efficient company prediction, we compared it with logit analysis, and discriminant analysis. The experimental results show that SVM provides a promising alternative for the prediction of an efficient company.

Keywords: IT Venture Business; Data Envelopment Analysis (DEA); Data Mining

Introduction

One of the fastest growing business fields has been the IT venture businesses for the last decade in Korea. Venture business in the IT industry is defined as a small company of high risk and high revenue. It is fast growing, and research based with an emerging technology. Many Korean venture businesses started appearing in 2000, that was the so called "IT bubble". While a lot of companies went bankrupt or were on the verge of bankruptcy, the companies which overcame the fierce competition for survival have grown explosively. Achieving a high market share is paramount importance in any emerging industry. In addition, cost control, profitability, income, and productivity are also important. IT venture business has to stand by cutting down their expenditure and improving their productivity until they gain revenue enough to survive the competitive market. Many IT venture business failed to convert the initial expenditure to actual income eventually. Investors or loan mangers should identify the venture business ensuring a large profit or promised interest as well as their invested capital or loaned money. However, no one knows if the IT venture business will succeed or fail in achieving a minimum market share for ongoing business.

In Korea, the number of IT venture businesses in the area of software including telecommunications and games have been unprofitable during several years. Although adequate expenditure is very important in an emerging industry, there must be a limit to how much a firm can spend in order to establish itself in the marketplace. Because of their inefficiency and inability to convert their initial expenditure into a positive income most inefficient companies failed in the year 2000 (Min and Lee, 2005). Some companies have been listed on the KODAQ, which is the stock market division of the Korean exchange for dealing with the securities of the venture business. So investors around KOSDAQ have tried to measure and analyze the efficiency of IT venture businesses eagerly in light of the efficient operations of technologies and utilization of assets. It is very useful for them to find efficient companies, with little expenditure, which still manage to obtain high levels of income. Because achieving a high income is a very important element in any emerging industry the IT venture businesses have been no exception. Therefore, it is necessary to evaluate the efficiency of the firms and to understand the important factors which determine their efficiency.

There have been two popular approaches to resolve the assessment of IT venture business. One approach uses financial information by employing a parametric model. Since Altman (1968) introduced this approach, a lot of studies have shown that statistical and artificial techniques such as multidiscriminant analysis, logit, neural networks, and support vector machines are outstanding tools for predicting the success or the failure of business in future. The only and decisive weakness is that it requires enough financial information for building the model because IT venture business is too young to collect the financial information for building a parametric model. Another approach utilizes nonfinancial information such as the possibility of growth, stability, and profitability. The judgment for nonfinancial information depends on experts who are in charges of assessing IT venture business. Credit ratings for well-known companies are presented by the credit rating agency such as Moody's, Standard & Poor's, and Fitch. However, these agencies couldn't assess it whenever an investor or a loan manager needs to know the efficiency of a venture business. It stands to reason that the agency presents only a periodic or scheduled report including the assessment an IT venture business. To resolve these problems mentioned in the two approaches, this study proposed an approach which utilizes measuring the efficiency of IT venture business by non-parametric model such as DEA (Data Envelopment Analysis). Many studies measured the efficiency of companies using DEA (Barua et al., 2004; Carlos et al., 2005; Koo et al., 2006). The proposed approach predicts the efficiency of IT venture business from measured DEA efficiency by data mining techniques such as support vector machines.

In this study we developed a model for measuring and analyzing the efficiency of an IT venture business. We employed the DEA and data mining techniques for our model. Using the DEA model, we evaluated the efficiency of a specific company without need for comparison. Furthermore, using a Support Vector Machine (SVM), one of machine learning techniques, we are able to predict efficient companies without repeated analyzing.

In this paper, our research objectives are as follows:

One is to make a model for evaluating the efficiency of IT venture business using DEA efficiency. The other is to develop a parametric model for assessing of an IT venture business without DEA analyses, which means to build another DEA model repeatedly whenever an added firm is assessed.

Literature Review

Data Envelopment Analysis

The DEA (Data Envelopment Analysis) is a typical approach among non-parametric methods for measuring efficiency, and was employed as our model. The DEA is a non-parametric multiple input-output efficiency technique that measures the relative efficiency of decision making units (DMUs) using a linear programming based model. It is non-parametric because it requires no assumption on the shape or parameters of the underlying production function. Traditional DEA is based on the pioneering work of Farrell's efficiency measure (1957). Then, Charnes et al. (1978) introduced CCR measures, and Banker et al. (1984) developed BCC measures. In the DEA model, the efficiency is calculated to multiple inputs and outputs. Two models, which are well-known as the most representative in DEA efficiency measures, are the CCR and the BCC model.

The CCR model is calculated with the constant return to scale (CRS) assumption. Banker et al.(1984) presented the BCC model, which allows for a variable return to scale and computes the scores of technical efficiency (TE) and scale efficiency (SE) for each firm in a data set. In these models, the DMUs were not penalized for operating at a non-optimal scale. These models yield the same results if achievement of efficiency or failure is the only topic of concern (Cooper et al., 2000).

DEA results are classified as either an efficient group or an inefficient group. DEA yields projections of inefficient DMUs onto an efficient piecewise linear frontier (Golany and Roll, 1989). Although DEA is a powerful tool for the efficiency measurement, there are some important things that have to be considered. One of them is the selection of inputs and outputs. It is important to select appropriate inputs and outputs, because combinations of selected inputs and outputs generate different efficiency rankings for DMUs.

DEA has been applied for evaluating the relative efficiency of DMUs in many places such as hospitals, schools, banks, factories, and retail stores. Particularly, DEA is applied to the evaluation of company and corporate credit ratings (Lee, 2006).

In prior research, we found studies applying DEA to the evaluation of a company. Lee et al. (2007), Carlos et al.(2005) and Barua et al. (2004) applied DEA to measure the efficiency of Internet companies. Lee et al. (2007) used the four inputs of capital, assets, salary, and advertising expenditure. The two outputs were visitors and sales. Carlos et al. (2005) applied the three inputs and the two outputs. The inputs were employees, expenses and assets, and the outputs were visitors and revenues. Barua et al. (2004) employed IT capital, NIT capital, labor, and number of years in business for inputs, and they used sales and gross margins for outputs. Koo et al. (2006) and Kim (2004) researched KOSDAQ companies by using the DEA. Koo et al. (2006) used total assets, employees, cost of sales, selling and administrative expenses for inputs, and sales for output. Kim (2004) used total assets, employees, and cost of sales for inputs and sales for the output in his study. The review of research shows that many researchers are applying DEA to evaluate companies. In this study, DEA was employed for the evaluation of the efficiency of an IT venture business.

Support Vector Machine

The Support Vector Machine (SVM) is a popular technique for solving data classification problems. We employed SVM to predict the efficiency in IT venture companies. The SVM method was developed by Vapnik (1995) SVM, one of many machine learning techniques, is based on statistical theory. It has shown good performance and a generalizing capacity in classification tasks. It is applied to the many areas of business (Tay and Cao, 2001).

SVM is the algorithm that finds the maximum margin hyperplane, which is the maximum separation between classes. In here, support vectors are the closest to the maximum margin hyperplane. If it is impossible to divide into two classes, we can use the kernel function. In the case of nonlinear class boundaries, we can transform the inputs into a high-dimensional feature space. This is the original input space and is mapped into a high-dimensional dot-product space.

In the separating case, we can presume the function $f : R^n \to \{\pm 1\}$ using a training set. In the separated two classes, A is defined as $x_i \in R^n$, $y_i = 1$, B is defined as $x_i \in R^n$, $y_i = -1$. If it is possible to separate them linearly, they can be represented in equations (1) and (2).

$$w \cdot x_i + b \geq +1, \ ^\forall x_i \in A \qquad (1)$$

$$w \cdot x_i + b \leq -1, \ ^\forall x_i \in B \qquad (2)$$

Where x is the input vector, w is the weight vector and b is bias. w and b represent the parameters used to determine the hyperplane.

Using equations (1) and (2), we can derive equation (3), as follows:

$$y_i (w \cdot x_i + b) \geq 1, \ ^\forall x_i \in A \cup B \qquad (3)$$

The maximum margin classifier optimizes data within the maximum margin hyperplane. This is an optimization problem expressed equation (4):

$$\min_{w,b} \ \frac{w \cdot w}{2} \qquad (4)$$
$$s.t. \quad y_i (w \cdot x_i + b) \geq 1$$

Finally, the equation for an optimal separating hyperplane is shown in equation (5).

$$f(x, \alpha_i, b) = \sum y_i \alpha_i (x \cdot x_i) + b \qquad (5)$$

Where α_i and b are parameters for determining the separation of the hyperplane. x is the training data, and x_i is the support vector.

In the case of nonlinear class boundaries, we can implement the idea by transforming the inputs into the high-dimensional feature space. A nonlinear separating case, is represented in equation (6).

$$f(x, \alpha_i, b) = \sum y_i \alpha_i K(x, x_i) + b \qquad (6)$$

Where $K(x, x_i)$ is called the kernel function. The examples of the kernel functions are the polynomial kernel $K(x, x_i) = (x \cdot x_i + 1)^d$, and the Gaussian radial basis function $K(x, x_i) = \exp(-\frac{1}{\delta^2}(x - x_i)^2)$.

Lee (2007), Huang et al. (2007), Gestel et al. (2004), Chen and Shih (2006) and Huang et al. (2004) applied SVM to the problem of estimating credit rating. Wu et al.(2007), Min and Lee (2005), Shin et al. (2005) and Min and Lee (2005) studied the prediction of bankruptcy using SVM. In addition, Tay and Cao (2002) used SVM to forecast a financial time series, and Hao et al. (2007) employed SVM in categorizing the document. Like this, SVM is applied to many areas for the reliable prediction of performance

Research Framework

In this study, we proposed an integrated method using DEA and data mining techniques to evaluate and predict the efficiency of an IT venture business. Figure 1 shows our research framework. To measure the efficiency of IT venture businesses, we collected the corporate information. We are able to obtain this information from IR reports or the corporate information provider companies. Generally, the corporation information contains information about an outline of a business, a value of company, and welfare of workers. An outline of a business includes the name of the firm, a representative, a main office or branch offices. The quantitative items such as a statement of profit and loss and a balance sheet, safety, or potential represent a value of a company. Welfare of workers includes pay and benefit package. Applying IT venture businesses to the DEA, we evaluated a relative efficiency of them.

Experiments and Results

The objective of this section is to analyze the efficiency of an IT venture business. This section starts by describing the data collection. Then it continues with a discussion of the model specifications, which includes the selection of inputs and outputs that enter the model. It ends by producing DEA efficiency estimates and analyses.

Data

In this study, we chose 93 KOSDAQ companies associated with software in the IT venture business. Among them, we used the information of 71 companies, excepting the companies for which information was insufficient. We obtained company information from 2005 from the KIS (Korea Information Service, Inc.).

Figure 1. The framework of DEA-Based data mining
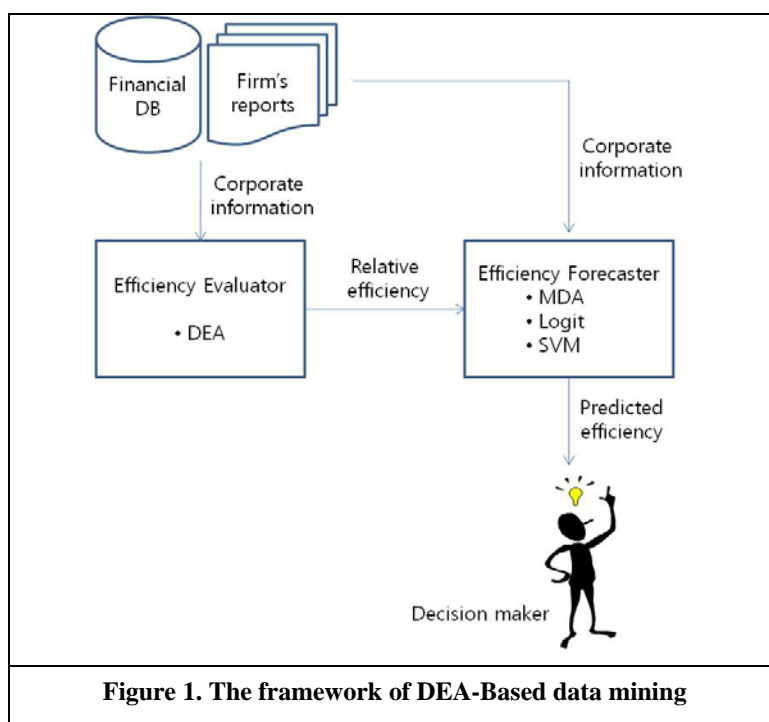
Model Specification and Results

The 71 IT venture businesses were used for our DEA programming. In DEA modeling, the selection of inputs and outputs is very important because selected inputs and outputs generate different efficiency rankings of DMUs. In this study, we employed four inputs. They were the number of employees, total assets, development cost, and selling and general administrative expenses. Sale was used for the output. The mean, standard deviation, and maximum and minimum were calculated for the input and output variables. The results can be seen in table 1.

The efficiency and ranking results of the DEA are listed in the appendix 1. These results represented the characteristics of each company. The CCR results showed the fact that 39 out of 71 companies were below the average, assuming the constant returns-to-scale of the DEA model. Seven of the 71 DMUs were efficient companies, as they had DMUs of 3, 5, 25, 31, 46, 47 and 54. These companies were more frequently referenced when evaluating inefficient companies. If the CCR efficiency is not equal to 1, we have to examine whether the reason is a managerial problem, or a scale problem, especially when the efficiencies of DMUs of 7, 12 and 33 are under 0.2. These companies had high input variables, but sales which were relatively low. In these cases, there are managerial problems.

Table 1. The descriptive statistics (Unit: the number of person, mil. Won)					
	Variables	Mean	S.D[1]	Max.	Min.
Input	Employees	106.44	10.452	679	21
	Total asset	37,625	3,841	203,220	7,598
	Development cost	1,841	378	21,112	3
	Selling & General admin. expenses	6,452	664	35,840	742
Output	Sale	18,646	1,407	53,558	3,600

1) The standard deviation

The BCC scores provide efficiency evaluations using a local measure of scale, i.e. under the variable return-to-scale. In this model, DMUs of 11, 18, 22, 30, 49, 58, 59, 66 and 67 were accorded efficient status in addition to the 7 CCR efficient companies which retained their previous efficient status. The DMU's full efficiency with the BCC model was due to its use of the smallest amount of inputs even though it had the lowest CCR score. In the BCC model, the companies exhibited scores in which 32 out of 71 were below average. The scale efficiency as defined by the ratio, CCR efficiency/BCC efficiency. For example, DMU 1 had a low BCC score and relatively high scale efficiency among the group, meaning that the overall inefficiency in the CCR column of DMU 1 was caused by inefficient operations rather than scale inefficiency. DMU 11 had a fully efficient BCC score and low scale efficiency. This can be interpreted to mean that the global inefficiency of this company under is mainly attributed to disadvantageous scale conditions. In the score of the return to scale, as identified by the BCC model, companies with full efficiency in the CCR score were also efficient in the BCC model, the region where constant returns-to-scale prevails. DMU 3, 5, 25, 31, 46, 47 and 54 had this status, while all other companies displayed increasing return-to-scale.

DEA-Based Data Mining

We sorted relatively efficient companies by using the DEA methodology. This method is useful for distinguishing relatively efficient and inefficient companies. However, it cannot be used to evaluate a company's absolute efficiency. Therefore, in this step, efficiency scores were applied to a logit model. First of all, we performed a paired t-test and then a logit analysis using significant financial variables.

Logit analysis has been utilized to investigate the relationship between binary or ordinal response probabilities and explanatory variables. The method fits the linear logit model for binary or ordinal response data via the method of maximum likelihood (Hua et al. 2007). In our study, we divided IT venture businesses into the two classes according to their ranking in the CCR efficiency scores. The two classes were efficient companies and inefficient companies. Using the stepwise module in SPSS 12.0, we conducted the logit analysis.

Table 2 presents the result of the logit model at the 0.5 cutoff. The results from, the prediction probability of the classification accuracy is 88.2% for inefficient companies. In the case of the efficient companies, the prediction probability was 83.8%. Using this criterion, it correctly classified 85.9% of IT venture businesses.

Table 2. The result of logit analysis			
Observation	Prediction		
	Efficiency	Inefficiency	Classification Accuracy (%)
Efficiency	31/37	6/37	83.8
Inefficiency	4/34	30/34	88.2
Total (%)			85.9

Table 3 shows the information for the financial variables included in the regression formula. The coefficient estimate of the total capital turnover was significant with a $p < 0.01$. The coefficient estimate of the sales/employee

was significant with a p < 0.05. This model was found to be highly significant. We know that these two financial variables are good criteria for determining an efficient company.

Table 3. The results of logit analysis				
Variables	B[1]	S.E.[2]	Wald[3]	Sig. Prob.[4]
Total Capital Turnover	5.818	1.696	11.771	0.001**
Sales/Employee	0.000	0.000	4.581	0.032*
Constant	-5.990	1.475	16.490	0.000**

1) Coefficient estimates, 2) standard error, 3) Wald statistics,

4) Significant probability, * p<0.05   ** p<0.01

In this study, the Gaussian radial basis function was used as the kernel function of SVM. In SVM, the upper bound C and the kernel parameter δ^2 play an important role. C and δ^2 have a big effect on the performance of SVM. In this study, we used LIBSVM-2.83(Chang and Lin, 2006). For the SVM experiment, we applied 20, 40, 60, 80, and 100 to C. The value of δ^2 was set at 1, 25, 50, 75, and 100.

When δ^2 was 25 in datasets 1, 2, 3, a better prediction performance was presented. In that case, the value of C was 60, 20 and 20 for each dataset. And when δ^2 was 50 in dataset 5, a better prediction performance was recorded. In that case, C was 60. In dataset 4, a higher prediction performance was found when δ^2 was 100 and C was 20.

Comparing Performance

In this study, we divided data into a training data set and a test data set. Then we performed a 5-fold cross validation with 70 pieces of data to verify the findings of SVM, logit, and discriminant analysis. Table 4 shows the prediction results of the efficient and inefficient companies.

Table 4. The prediction results (unit: %)						
	Discriminant		logit		SVM	
	Train	Test	Train	Test	Train	Test
Set 1	82.1	78.6	83.9	85.7	80.4	85.7
Set 2	78.6	92.9	82.1	92.9	78.6	92.9
Set 3	87.5	78.6	85.7	71.4	82.4	78.6
Set 4	83.9	78.6	85.7	78.6	82.1	78.6
Set 5	82.1	78.6	85.7	85.7	85.7	85.7
average	82.8	81.5	84.6	82.9	81.8	84.3

We compared results of SVM, logit, and discriminant analysis. The hit ratio results show that SVM provides a promising alternative for the prediction of efficient companies.

Conclusions

We proposed a DEA-Based approach to evaluate efficiency in an IT venture business and performed an empirical analysis for the companies listed on the KOSDAQ. In this study, we used data mining, to discover advantageous patterns in data for the prediction of an efficient IT venture business. We divided the companies into two groups in

accordance with the efficiency scores of the DEA model. Using the logit model through the stepwise method, we finally acquired a model for evaluating the efficiency of an IT venture business. In the case of analyzing a new company, the DEA model demands other companies to be compare with the new company to generate the relative efficiency. Our integrated model could evaluate the efficiency of a specific company. Through this process, we could find the financial variables which characterize an efficient company. Further through the application SVM, we could predict efficient companies.

As a result, we applied our integrated model to companies listed on the KOSDAQ, with corporate information available from 2005. Our model enabled us to evaluate an individual firm and provide the efficiency information of an IT venture business without comparing it with other companies. For IT venture businesses in Korea, we found aspects, such as a total capital turnover, sales/employee, and the productivity of employees to be very important pieces of financial information to evaluate the efficiency of an IT venture business. Also, to examine the feasibility of SVM in efficient company prediction, we compared performances of logit analysis, and discriminant analysis. The experimental results show that SVM provides a promising alternative for efficient company prediction.

We should generalize our integrating model by applying it to the IT venture businesses of other countries. Through this additional analysis, we will be able to better understand the general features of the IT venture business.

References

Altman, E. 1968. "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," *The Journal of Finance* (23), pp.589-609.

Banker, R.D., Charnes, A. and Cooper, W.W. 1984. "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis," *Management Science* (30:9), pp. 1078-1092.

Barua, P.L., Brockette, Cooper, W.W., Deng, H., Parket, B.R., Ruefli, T.W. and Winston, A. 2004. "DEA Evaluations of Long- and Short-Run Efficiencies of Digital vs. Physical product "Dot Com" Companies," *Socio-Economic Planning Sciences* (38:4), pp. 233-253.

Carlos, S.C., Yolnada, R.C. and Cecilio, M.M. 2005. "Measuring DEA efficiency in Internet Companies," *Decision Support Systems*, (38:4), pp. 557-573.

Chang, C.-C., and Lin, C.-J. 2006, LIBSVM: a Library for Support Vector Machines, http://www.csie.ntu.edu.tw/~cjlin.

Charnes, A., Cooper, W.W. and Rhodes, E. 1978. "Measuring the Efficiency of Decision Making Units," *European Journal of Operational Research* (2:6), pp. 429-444.

Chen, W.H. and Shih, J.Y. 2006. "A study of Taiwan's Issuer Credit Rating Systems Using Support Vector Machines," *Expert Systems with Applications* (30:3), 427-435.

Cooper, W.W., Seiford, L.M. and Tone, K. 2000. *Data Envelopment Analysis*, Kluwer Academic Publishers, Norwell, MA, 188-191.

Farrell, M.J. 1957. "The Measurement of Productive Efficiency," *Journal of the Royal Statistical Society* (120), pp.253-281.

Gestel, T.V., Baesens, B., Dijcje, P.V., Garcia, J., Suykens, J.A.K. and Vanthienen, J. 2004. "A Process Model to Develop an Internal Rating System: Sovereign Credit Ratings," *Decision Support Systems* (42: 2), pp. 1131-1151.

Golany, B. and Roll, Y. 1989. "An Application Procedure for DEA," *Omega* (17:3), pp. 237-250.

Hao, P.Y., Chiang, J.H. and Tu, Y.K. 2007. "Hierarchically SVM Classification Based on Support Vector Clustering Method and Its Application to Document Categorization," *Expert Systems with Application* (33 :3), pp. 627-635.

Huang, C.L., Chen, M.C. and Wang, C.J. 2007. "Credit Scoring with a Data Mining Approach Based on Support Vector Machines," *Expert Systems with Application*, (33:4), pp. 847-856.

Huang, Z., Chen, H., Hsu, C.J., Chen, W.H. and Wu, S. 2004. "Credit Rating Analysis with Support Vector Machines and Neural Networks: a Market Comparative Study," *Decision Support Systems*, (30:4), pp. 543-558.

Kim, D.H., Oh, S.B. and Choi, Y.H. 2005. "An Empirical Study on the Determinants of Stock Price in KOSDAQ Venture Firms of KOREA," *in Proceedings of the 3rd International Joint Conference among KALM MMDIC and MECS*, pp. 98-107.

Kim, J.J. 2004. "The Effect of DEA Efficiency on ROE of KOSDAQ Firms," *The Research of Tax and Accounting* (14), pp. 153-170.

Koo, C.O., Kim, G.J. and Park, J.W. 2006. "The Productivity Analysis of KOSDAQ Software Companies," *Journal of Accounting & Finance (*6:1), pp. 27-51.

Lee, C.W. 2005. "A Study on the Concept and Essential of Business Venture," *Business Research* (34:2), pp. 315-348.

Lee, K.J., Kim, J.J. and Cho, G. 2007. "The Evaluation of the Efficiency of Internet Companies Using a DEA Model," *Business Research* (19:5), pp. 109-136.

Lee, Y.C. 2006. "Credit Risk Evaluation Using Worst Practice DEA," *Daehan Journal of Business* (19:5), pp. 1711-1735.

Lee, Y.C. 2007. "Application of Support Vector Machines to Corporate Credit Rating Prediction," *Expert Systems with Application* (33:1), pp. 67-74.

Min, J.H. and Lee, Y.C. 2005. "Bankruptcy Prediction Using Support Vector Machine with Optimal Choice of Kernel Function Parameters," *Expert Systems with Application* (28:4), pp. 603-614.

Min, S.H., Lee, J. and Han, I. 2006. "Hybrid Genetic Algorithms and Support Vector Machines for Bankruptcy Prediction," *Expert Systems with Application* (31:3), pp. 652-660.

Shin, K.S., Lee, T.S. and Kim, H.J. 2005. "An Application of Support Vector Machines in Bankruptcy Prediction model," *Expert Systems with Application* (28:1), pp. 127-135.

Tay, F.E.H. and Cao, L. 2001. "Application of Support Vector Machines in Financial Time Series Forecasting," *Omega*, (29:4), pp. 309-317.

Tay, F.E.H. and Cao, L.J. 2002. "Modified Support Vector Machines in Financial Time Series Forecasting," *Neurocomputing*, (48), pp. 847-861.

Vapnik, V. 1995. *The Nature of Statistical Learning Theory*, Springer.

Wu, C.H., Tzeng, G.H., Goo, Y.J. and Fang, W.C. 2007. "A real-valued Genetic Algorithm to Optimize the Parameters of Support Vector Machine for Predicting Bankruptcy," *Expert Systems with Application*, (32 :2), pp. 397-408.

Yu, S.H. and Park, M.C. 2000. "A Study on Marketing Strategies of IT Ventures with Venture Performance Model," *in Proceedings of the Conference on the Korean Operations and Management Science Society*, pp. 293-297.

Appendix

DMU No.	CCR		BCC		S.E[1]	RTS[2]	# of ref.	
	efficiency	ranking	efficiency	ranking			CCR	BCC
1	0.4747	41	0.5072	58	0.9359	Increasing	0	0
2	0.5868	28	0.8397	25	0.6988	Increasing	0	0
3	1.0000	1	1.0000	1	1.0000	Constant	26	8
4	0.4002	51	0.5588	50	0.7162	Increasing	0	0
5	1.0000	1	1.0000	1	1.0000	Constant	23	10
6	0.8294	10	0.8365	26	0.9915	Decreasing	0	0
7	0.1919	66	0.4616	63	0.4158	Increasing	0	0
8	0.3443	55	0.8553	23	0.4026	Increasing	0	0
9	0.5053	37	0.7970	32	0.6340	Increasing	0	0
10	0.4212	49	0.5345	56	0.7880	Increasing	0	0
11	0.6497	24	1.0000	1	0.6497	Increasing	0	30
12	0.1760	68	0.3728	67	0.4720	Increasing	0	0
13	0.3803	53	0.5074	57	0.7494	Increasing	0	0
14	0.4704	42	0.5754	47	0.8176	Decreasing	0	0
15	0.2064	63	0.4803	61	0.4297	Increasing	0	0
16	0.7027	18	0.7036	39	0.9987	Decreasing	0	0
17	0.2732	59	0.2909	68	0.9391	Increasing	0	0
18	0.3806	52	1.0000	1	0.3806	Increasing	0	11
19	0.3002	58	0.4390	64	0.6839	Increasing	0	0
20	0.5301	36	0.5386	54	0.9843	Decreasing	0	0
21	0.8006	13	0.8492	24	0.9428	Increasing	0	0
22	0.6656	22	1.0000	1	0.6656	Decreasing	0	2
23	0.1964	65	0.8301	27	0.2366	Increasing	0	0
24	0.3039	57	0.5546	52	0.5480	Increasing	0	0
25	1.0000	1	1.0000	1	1.0000	Constant	24	28
26	0.8007	12	0.8262	28	0.9691	Decreasing	0	0
27	0.4478	48	0.7692	34	0.5821	Increasing	0	0
28	0.5400	34	0.5559	51	0.9714	Decreasing	0	0
29	0.5771	31	0.8233	29	0.7010	Increasing	0	0
30	0.6923	19	1.0000	1	0.6923	Increasing	0	6
31	1.0000	1	1.0000	1	1.0000	Constant	59	36
32	0.4151	50	0.4963	59	0.8364	Increasing	0	0
33	0.1066	71	0.1071	71	0.9951	Increasing	0	0
34	0.7130	17	0.8012	31	0.8899	Increasing	0	0
35	0.4657	43	0.5654	49	0.8236	Increasing	0	0
36	0.4902	38	0.6904	40	0.7101	Increasing	0	0
37	0.4837	39	0.5865	46	0.8247	Increasing	0	0
38	0.4613	45	0.5510	53	0.8372	Increasing	0	0
39	0.8834	9	0.9308	19	0.9491	Increasing	0	0
40	0.5845	30	0.6142	45	0.9517	Increasing	0	0
41	0.7859	14	0.8824	21	0.8906	Increasing	0	0
42	0.4583	46	0.5357	55	0.8555	Increasing	0	0
43	0.6793	20	0.9098	20	0.7467	Increasing	0	0
44	0.7156	16	0.7325	36	0.9769	Decreasing	0	0
45	0.1893	67	0.2247	70	0.8423	Increasing	0	0
46	1.0000	1	1.0000	1	1.0000	Constant	8	5
47	1.0000	1	1.0000	1	1.0000	Constant	9	7
48	0.6554	23	0.7295	37	0.8984	Increasing	0	0
49	0.9593	8	1.0000	1	0.9593	Increasing	0	11
50	0.2453	61	0.3932	65	0.6238	Increasing	0	0

DMU	CCR		BCC		S.E[1]	RTS[2]	# of ref.	
No.	efficiency	ranking	efficiency	ranking			CCR	BCC
51	0.4615	44	0.6434	43	0.7174	Increasing	0	0
52	0.5858	29	0.9487	18	0.6175	Increasing	0	0
53	0.7575	15	0.7960	33	0.9517	Increasing	0	0
54	1.0000	1	1.0000	1	1.0000	Constant	5	4
55	0.4803	40	0.4816	60	0.9972	Decreasing	0	0
56	0.6082	27	0.6815	41	0.8925	Increasing	0	0
57	0.1291	70	0.2255	69	0.5726	Increasing	0	0
58	0.3321	56	1.0000	1	0.3321	Increasing	0	1
59	0.6289	26	1.0000	1	0.6289	Increasing	0	26
60	0.1973	64	0.6427	44	0.3070	Increasing	0	0
61	0.1412	69	0.6790	42	0.2080	Increasing	0	0
62	0.4515	47	0.4641	62	0.9728	Increasing	0	0
63	0.2174	62	0.5700	48	0.3813	Increasing	0	0
64	0.8223	11	0.8652	22	0.9504	Increasing	0	0
65	0.2590	60	0.3910	66	0.6624	Increasing	0	0
66	0.5613	32	1.0000	1	0.5613	Increasing	0	12
67	0.5373	35	1.0000	1	0.5373	Increasing	0	6
68	0.6414	25	0.7206	38	0.8901	Increasing	0	0
69	0.3601	54	0.7672	35	0.4694	Increasing	0	0
70	0.5443	33	0.9683	17	0.5621	Increasing	0	0
71	0.6716	21	0.8167	30	0.8223	Increasing	0	0
average	0.5426		0.7172		0.7555			
S.D	0.2481		0.2313		0.2192			
# effic.	7		16		7			

1) standard error, 2) return to scale