Nonlinear Models and Linear Models in Expert-Modeling : A Lens Model Analysis

Choong Nyoung Kim

전문가 모델링에서 비선형모형과 선형모형 : 렌즈모형분석

김 충 녕

ABSTRACT

The field of human judgment and decision making provides useful methodologies for examining the human decision making process and substantive results. One of the methodologies is a lens model analysis which can examine valid nonlinearity in the human decision making process. Using this method, valid nonlinearity in human decision behavior can be successfully detected.

Two linear (statistical) models of human experts and two nonlinear models of human experts are compared in terms of predictive accuracy (predictive validity). The results indicate that nonlinear models can capture factors (valid nonlinearity) that contribute to the experts' predictive accuracy, but not factors (inconsistency) that detract from their predictive accuracy. Then, it is argued that nonlinear models can be more accurate than linear models, or as accurate as human experts, especially when human experts employ valid nonlinear strategies in decision making.

1. Introduction

Human decision-making has been studied for a long time, normatively or descriptively. One of the major

* Department of Business Administration School of Social Science Hallym University

research issues is to construct descriptive decision models which describe, mimic, and replace human judgment. In attempting to construct the models of judgment, there are two dominant paradigms; output analysis and process-tracing. The former paradigm focuses on finding relationships between the relevant cues and final decisions of humans, and developing a decision model simulating these relationships. The latter paradigm focuses on analyzing the actual process of making judgment, and, as a result, developed various methods for the analysis of the process of making judgment.

With the advancement of computer technology, the judgment-modeling issue has been revisited in expert system research, with a different name, expert-modeling. How to extract knowledge from an expert to build a reliable knowledge base is one of it's concerns. As computer scientists pursued the same goal as the output analysis approach, the valuable methodologies of the process-tracing approach have never been coupled with expert-modeling research.

This paper applies one of the famous process-tracing methods, lens-model analysis, to expert-modeling, by 1) analyzing valid nonlinearity in decision process, 2) comparing predictive validity between two linear models and two famous inductive learning methods as nonlinear models, and 3) examining the contingent relationship between the valid nonlinearity and predictive validity of the four models.

2. Literature Review

This section reviews key literature in the process-tracing approach and empirical studies in expert-modeling as an extension of judgment-modeling. Two research questions are also raised based on the literature review.

2.1. Lens Model Studies

First proposed by Brunswick [1], the lens model has been described by others [11][26] to investigate the use of nonlinearity in human decision-making behavior. Tucker [26] showed that five indices are functionally related in a general equation :

$$ra = G{*}R_t{*}R_f + C(1-R_t^2)^{1/2}(1-R_f^2)^{1/2} \tag{1}$$

where : C : correlation coefficient of residuals corresponding to Y_t and Y_f

Y_e : actual (normative) outcomes in environmental data base,

Y_s : human expert's decisions,

Y_t : predicted values (outcomes) from a linear regression ofactual outcomes on cues,

Y_f : predicted values (decisions) from a linear regression ofdecisions of human expert on cues,

r_a : correlation of Ys and Y_e,

G : correlation of Yt and Y_f,

R_t : correlation of Y_e and Y_t, and

R_f : correlation of Y_s and Y_f.

The index C is defined as the correlation between the variance which is not accounted for by the environmental regression model and the variance which is not accounted for by the human expert's regression model. The index can also be considered to be the partial correlation coefficient of Y_e and Y_s when the linear effects of all the cues are totally removed [15]. This index can take on values between 1.00 and −1.00. Many behavioral accounting researchers have discussed the use of the lens model with regard to the examination of a judgment situation in which a human makes decisions [17].

Therefore, the residual index (C) can be used as a general measure of nonlinearity in the decision behavior of a human expert. If either of the residual variances of the two linear regression models is random or there is no systematic relationship between them, the correlation coefficient C approaches zero (insignificant level of nonlinearity). This indicates that the decision strategies being used by experts are close to linear and that the decision making behavior of a human expert can be modeled well by a linear or compensatory model.

If C shows positive correlation, there are significant levels of nonlinear components in the decision strategies of the expert. Thus, the expert uses valid nonlinear strategies during his/her decision-making process which cannot be captured by a linear model. An inference can be made in this case : if nonlinear algorithms areapplied to capture the valid strategies, more accurate decision models can be built. If C shows negative correlation, then the nonlinear strategies being used by the experts seem to be deteriorating their predictive accuracy.

2.2. Output-analysis (Bootstrapping)

Output analysis is also called bootstrapping in accounting literature. Bootstrapping is the replacement of a decision maker by the simple linear model of his or her judgment. Linear (Logistic) regression and discriminant analysis (DA) have been used primarily to build the linear models. Many studies [10, 27, 2] provide evidence on the applicability of bootstrapping, supporting the strength of the linear models. The results of those studies show that the linear models of human subjects outperform the human subjects in predicting the actual outcomes in environment.

However, some others [16] contradicted the possibility of bootstrapping because human decision behavior basically follows nonlinear strategy and it can never be modeled well by linear models. What the linear models can do is just approximate the decision behavior. Einhorn et al. [8] and Larker and Lessig [14] proved in their protocol analysis that human decision making follows nonlinear strategies, although these studies did not exclude the possibility of bootstrapping by linear models.

Libby [16] also argued that, as the quality of human knowledge improves, nonlinearity of human decision behavior becomes obvious and difficult to be modeled by linear models. In other words, if a real expert with a high quality of knowledge is to be modeled by a linear model, the valid nonlinearity of the expert should be approximated. It means that the linear model may not appropriately represent the valid nonlinearity, and its performance (predictive validity) may deteriorate and be worse than the performance of the expert. To support his argument, Libby [16] reviewed the previous studies which reported successful bootstrapping by linear modelsand found that these studies did not use domain experts but novices as subjects. Non-experts tend to be more inconsistent and less valid than experts.

It is a very interesting issue in decision research if, in predicting the actual outcomes, the model of human expert's judgment would be more valid or less valid than the expert's judgment [16][15]. To the extent that the model fails to capture the valid nonlinear decision strategy of an expert, it should perform worse than the expert. To the extent that the model eliminates the inconsistency component in the expert's decision-making behavior, it should outperform the expert. Therefore, the success of modeling the judgment of an expert seems to depend on how much a model can capture the valid nonlinear strategy of the expert as well as how much a model can eliminate the inconsistent component.

2.3 Expert-modeling and Inductive Learning (Machine Learning)

Expert-modeling means to use various mechanical algorithms to capture human expertise, knowledge and heuristic. This approach is also called inductive learning or machine learning, since it produces decision rules from the example cases a human expert provides [19]. A number of different inductive learning methods have been developed. Among them, neural network and ID3 approaches are most popularly used.

ID3 is considered a simple but effective rule-based method for learning by examples (LBE). When it is given a training set of positive and negative examples, ID3 constructs a decision tree for classifying examples into two classes [20].

The inductive learning method is based on the information theory and uses an information-theoretic measure called entropy. Entropy is a measure of the amount of information carried by a message in communication [24]. Based on the entropy concept, information on each attribute (independent variable) is processed sequentially and combined logically (conjunctively or disjunctively) along the decision tree [9]. Therefore, a linear combination of input information and other properties of linear models cannot be expected in the method.

Rather, the logical information processing in ID3 seems to be very similar to a nonlinear decision strategy. Thus, ID3 is considered a typical nonlinear algorithm.

Recently, another inductive learning algorithm was introduced : a neural network. A neural network is a dynamic model consisting of perceptrons (also called nodes), connections between the perceptrons, and layers associated with each perceptron [23]. There are three types of perceptrons : input, output, and hidden layer perceptrons. The input perceptrons receive input values from sources external to the neural network. The output perceptrons produce output of the neural network. The hidden layer perceptrons serve to detect features, regularities, and generalizations in the data. Hidden layers allow neural networks to perform more flexible information processing and make neural networks different from linear models [18].

Use of neural networks (another inductive mechanism) as a decision model has been addressed by a number of researchers [6, 5, 22, 25]. For example, Dutta and Shekhar [6] used a neural network to build a decision model and contrast it with a regression model in predicting bond-ratings. They found that neural nets outperformed linear models in a multiple classification task (bond-rating).

In spite of the algorithmical difference, the inductive learning algorithms were used to pursue the same goals as bootstrapping by linear models. For that reason, to validate models built by inductive learning algorithms, linear models such as linear regression and discriminant analysis were compared as a benchmark.

While performance of the inductive learning algorithms in general is good, some conflicting results [3][5] make it difficult to generalize the relative performance of inductive learning algorithms. Moreover, the results of most comparative studies tended to be dependent on data or task because they only compare the performance of two approaches without considering other exogenous factors than algorithms. Thus, the data-dependent tendency makes it more difficult to generalize the empirical results.

This study starts from a conjecture that in comparing the performance of models of human judgment, the characteristics of judgment behavior (for example, decision strategy and validity of the human judgment) should be analyzed. Then, a more reliable and comprehensive interpretation of the results can be achieved.

2.4. Research Objective

This study focuses on validity and nonlinearity of human decision making behavior. More specifically, linear and nonlinear strategies are differentiated and the validity of each strategy is also examined. To do this, 'C' index of lens model analysis is used, which is supposed to represent the valid nonlinearity in decision-making behavior.

The objective of this study is twofold. First, this study tries to prove the usefulness of the C index in expert-modeling by examining the relationship between the C index and the predictive validity of models. The high value of C index asserts the existence of valid nonlinearity in decision making behavior. Since the nonlinearity should be captured more successfully by nonlinear information-processing algorithms, its existence should contribute to the predictive validity of nonlinear algorithms.

Second, this paper examines Libby's argument [16] by analyzing the relationship between the predictive validity of the human expert to be modeled and the predictive validity of the models : 'The more valid (accurate) human subject is modeled, the less likely bootstrapping by linear models is because linear models cannot capture the valid nonlinear decision behavior of the human subject.'

Two hypotheses concerned with the objectives are as follows :

(1) There is a significant relationship between C index and the predictive validity of models with nonlinear information-processing capability.

(2) There is a significantly negative relationship between the predictive validity of the human expert to be modeled and the predictive validity of linear models : in other words, bootstrapping is not possible if the human subject is a real expert who has highly accurate predictive validity.

By testing these two hypotheses, we can examine if bootstrapping by linear models is possible all the time. Otherwise, the results of this study may support the use of nonlinear algorithms in a certain modeling situation.

3. Methodology

3.1. Subject

Eight loan officers from two commercial banks : a bank in California and a bank in Texas were contacted through senior administrators in their respective banks. Based on predictive validity and reliability, only three of the experts were selected and asked to complete the experiment. The average work experience in their positions was about ten years. In addition, a decision scheme was considered as the fourth expert : the composite expert. Decisions of the composite expert consisted of the majoritydecisions among the decisions on each case made by experts.

This experiment was conducted in three months. It began with interviews of the lending officers. Each interview was pre-arranged and was conducted individually in the office of each expert. In the instructions given to the experts, the objective of this experiment was explained as an investigation of the decision strategy in forming decisions about whether a firm would be bankrupt or would default on the payment of its debt within one year or two years from the date on which its financial informationwas prepared. The experts were informed that (1) all the cases in the sample were real ; (2) about half of them had actually been bankrupt or defaulted on the payment of debt ; (3) the cases were presented in a random order. The experts were also informed that all the financial information were derived from audited financial statements two years before the default or bankruptcy date.

3.2. Data Sets and Firms in the Sample

There were three groups of firms, which were categorized by the size of total assets. The first group was composed of 55 firms which had average total assets of about $344 million. The 55 cases included 28 failed and 27 nonfailed cases. A firm was classified as failed if it experienced bankruptcy, default, or was liquidated for the benefit of creditors within two years of the date of the financial statements.

The second group was composed of 60 firms which had averag total assets of about $44 million. These 60 cases included 30 failed and 30 nonfailed cases.

Data for bankrupt firms in the first and second groups were selected from listings in the Wall Street Journal Index for the years 1981-1985 and from listings of deleted firms (due to liquidation) in the Moody's Industrial Manual. Data for nonbankrupt firms in the two groups were obtained from the same sources for the same period as that of the corresponding bankrupt firms. These nonbankrupt firms were comparable with the bankrupt firms in terms of asset size and type of business. The financial information about each firm (case) was gathered from public sources such as Moody's Industrial Reports and Moody's OTC Manual.

The third group was composed of 59 small business firms which were mainly collected from actual loan cases of a large commercial bank in California. The 59 cases included 28 failed (rejected) and 31 nonfailed (approved) cases. In this group, a case was defined as 'nonfailed' if it was approved and later evaluated as a good loan. A case was classified as 'failed' if it was rejected at the outset, or initially approved but later evaluated as a bad loan, or the firm experienced bankruptcy. To make this group different from the other two, commercial loan cases of relatively small businesses were collected. The average loan amount of $318,000 was equivalent to 7% of the average total assets of firms in this group. Additionally, to make the third group comparable with the other two, commercial loan cases with sufficient financial information were collected and all the qualitative data of each loan were ignored.

In this field setting, it was assumed that the size of the total assets reflected the stability and credibility of a firm, and indirectly represented the default risk of a firm. The default risk as a control variable is appropriate for this naturalistic setting such as bankruptcy prediction. However, the primary purpose of using the control variable is to create different risky situations in which different decision strategies might be triggered in the decision making behavior of loan officers.

In order to facilitate comparison with the results of the previous studies, this research used about the same number of nonbankrupt (approved) as bankrupt (rejected) cases. To isolate the possible effect of economic condition from the experiments, this study collected data mostly from the period of 1981 through 1985 because, according to the report of U.S. Commerce Department (U.S. Commerce Bureau Analysis), the economic condition of this period was average. For the same purpose, utilities, transportation, and financial companies were excluded because these companies have different financial structures and environments [13].

3.3. Tasks

Each expert was provided with financial profiles of real but disguised industrial companies. The companies were represented by ten commonly used financial ratios computed from the firms' financial statements. The first five ratios were chosen through a factor analysis [16] and the remaining five ratios were the most commonly cited ratios for bankruptcy prediction in the risk analysis literature [13]. Selection of the ten ratios (cues) was also accepted by the expert lending officers during the first interview. They agreed that the ratios would provide sufficient quantitative information for bankruptcy prediction and would not cause any information overload. In addition, by presenting the financial profiles in ratios rather than real numbers, the research design of the three groups could avoid any possible effects from real numbers.

The experts evaluated the financial profiles of each firm and were asked to judge whether each of the firms in the sample would be bankrupt. Based on the judgment, each expert made a decision on each case in the sample : approve or reject. This is a binary classification task based on the experts' judgment, typical of a business environment.

To examine the experts' use of financial ratios in the loan evaluation task, the loan officers were asked to indicate the degree of importance of each ratio after they finished making decisions on all the sample

Table 3.1 Selection of Financial Ratios

(Percentage Weights)

	Expert 1	Expert 2	Expert 3
N.I. / T.A.	10	30	15
C.A. / Sales	10	20	15
C.A. / C.L.	20	30	20
Cash / T.A.	0	0	20
T.D. / T.A.	25	0	5
(C.A./C.L.)/T.A.	10	10	5
R.E. / T.A.	25	10	20

* N.I. : Net Income T.A. : Total Assets

C.A. : Current Assets C.L. : Current Liability

T.D. : Total Debt R.E. : Retained Earnings cases provided. Based on the degree (weight), a final set of cues (ratios) was determined for building decision models. These ratios are defined in Table 3.1.

3.4. Procedures

This study consisted of two steps : (1) examining the experts' decision strategies with the lens model and (2) building decision models and evaluating model performance in predicting the actual outcomes (predictive validity of models). In the first step, the experts were given all cases in the three groups without the actual outcome of each case and were asked to provide decisions on all the cases. In this way, two case sets on every case were available ; each case with actual outcomes and the same case with the expert's decisions. With the cases, various correlation coefficients between the variables on both sides of the lens model were computed, showing the valid nonlinear components in the individual decision making process.

In the second step, decision models using four different algorithms were built and their performance was evaluated in predicting the actual outcomes (predictive validity).

3.5. Model-Building

To generate rule-based nonlinear decision models, a commercial package (Rule Master) applying the ID3 method [21] was used. During model-building, a few crashed (inconsistent) cases were found and these cases were excluded. A decision of the rules was decided to be incorrect if the decision of the tree was undefined.

To generate network-based models, a commercial package (Neural Works Professional II) applying the Back-propagation paradigm of neural network was used. The configuration of the network used in this study was 3-layer networks (one hidden layer with 14 processing units). The learning coefficients of the generalized-delta rule were selected (0.9 and 0.6 for learning rate and momentum rate) because they were reported to yield fast learning [23].

Since this study was concerned with binary classification, only a single output processing element was needed. This configuration was applied for constructing all networks. Learning was completed after 100,000 iterations in most networks and about 400,000 iterations were in a few nets. The decision was stated as : output element $> 0.65 \Rightarrow$ nonfailed (accepted), output element $< 0.35 \Rightarrow$ failed (rejected).

If value of output element was between 0.35 and 0.65, the decision was considered undefined and then the undefined decision was counted as incorrect.

To construct linear discriminant analysis models (DA), the seven ratios in a numeric form were used. As in building neural nets, the original numeric data were used without transformation. The DA models were implemented using SAS.

Given that both the response and the criterion were dichotomous, SAS, a statistical package that includes the Logistic Regression method, was used to generate the logistic models. The decision of the logistic model classified as : the value of the dependent variable $\geq 0.5 \Rightarrow$ accepted, the value of the dependent variable $< 0.5 \Rightarrow$ rejected.

4. Results

Table 4.1 summarizes the predictive validity of the expert loan officers. The predictive accuracy of the experts in this study is about the same as in previous studies.

4.1. Analysis of Decision Strategy

Table 4.2 summarizes two key correlation coefficients of the lens model framework. Based on the coefficients in this table, the validity of nonlinear strategy used by each expert is determined. The C indices show that valid nonlinearity exists in the decision strategies of both expert 2 and expert 4 in Group 2, and in the strategies of expert 3 in Group 3.

Table 4.1 Predictive Validity of Loan Officers

	Group 1 (N=55)	Group 2 (N=60)	Group 3 (N=59)	Mean Accuracy
Expert 1	72.7%	75.0%	81.4%	76.4%
Expert 2	72.7%	76.7%	78.0%	75.8%
Expert 3	65.5%	73.3%	78.0%	72.3%
Expert 4	78.2%	83.3%	81.4%	81.0%
Mean of Expert 1, 2, 3	70.3%	75.0%	79.1%	

* Libby (1976) : 74% ; Zimmer (1980) : 77% ; present study : 75%

The achievement index (ra) is another indicator of each expert's predictive validity. The fourth expert has the best achievement as shown in Table 4.1.

Table 4.2 A Summary of Lens Analysis

	Group 1		Group 2		Group 3	
	Ra	C-index (Test-stat.)	Ra	C-index (Test-stat.)	Ra	C-index (Test-stat.)
Expert 1	0.467,	-0.057 (-0.424)	0.5 ,	0.212 (1.642)	0.627,	0 (0)
Expert 2	0.467,	0.066 (0.486)	0.582,	**0.363 (2.814)	0.57,	0.207 (1.588)
Expert 3	0.331,	-0.108 (-0.802)	0.495,	-0.002 (-0.02)	0.559,	**0.282 (2.169)
Expert 4	0.491,	0.128 (0.95)	0.7, (2.726)	**0.352 (0)	0.631,	0

** (Phi Test Statistics : Pcritical at 0.05=1.96)

4.2 Predictive Validity

Predictive validity is measured by evaluating model performance in modeling the decisions of the expert

Table 4.3 Predictive Validity of Experts, and Models of Experts

	Predictive Validity	Rank in Predictive Validity	Simulation Accuracy
Most accurate Expert	81.0%	(1)	---
Linear models of Most accurate Expert	LR Model : 78.6% DA Model : 77.5%	(4) (5)	LR Model : 73.5% DA Model : 78.1%
Nonlinear models of Most accurate Expert	ID3 Model : 80.9% NN Model : %80.4%	(2) (3)	ID3 Model : 70.1% NN Model : 76.5%
Expert' Average	76.3%	(9)	---
LR Models' Average	76.5%	(7)	72.5%
DA Models' Average	74.8%	(10)	77.0%
ID3 Models' Average	76.4%	(8)	68.5%
NN Models' Average	76.9%	(6)	78.7%

LR : Logistic Regression ; DA : Discriminant Analysis

ID3 : ID3 ; NN : Neural Network

and then predicting the actual outcomes. As in the previous studies, the following measurements are made : (a) The expert's validity : the number (in percent) of correct predictions of the actual outcomes made by the expert. (b) The validity of the model of the expert : the number (in percent) of correct predictions of the actual outcomes made by the model of the expert. (c) The incremental validity of the model of the expert over the expert's decision itself : the validity of the model of the expert less the validity of the expert.

In general, all the models generated by the four algorithms proved to predict the actual outcomes quite well. In the previous studies [4], the validity of the linear models proved to be higher than that of human judgment in most cases. However, in the present study, there was no significant difference in mean validity between the human expert and the linear models, not even between linear models and nonlinear models. Human experts slightly outperform DA models. Moreover, when the most accurate expert was modeled, two linear models were surpassed in the predictive validity by the expert as well as by two nonlinear models. This validity is summarized in Table 4.3.

To provide insight into analysis, two lens model components and the validity of the models in the twelve experiments were combined to compute correlations among the variables. The variables are : (1) validity of the expert at each experiment (VE), (2) correlation between the decisions of the experts and the actual outcomes (ra), (3) residual correlation as an index of valid nonlinearity (C), (4) validity of Logistic Regression (VLR), (5) validity of discriminant analysis (VDA), (6) validity of ID3 (VID3), (7) validity of neural network (VNN), (8) incremental validity of Logistic Regression over the validity of an expert (INVLR) (which is computed by subtracting the validity of an expert from the validity of the model of the expert), (9) incremental validity of discriminant analysis over the validity of an expert (INVDA), (10) incremental validity of ID3 (INVID3) over the validity of an expert, (11) incremental validity of neural network over the validity of an expert (INVNN).

The correlations among the eleven variables are presented in Table 4.4. The significant correlations at the 0.05 or 0.1 significance level) were (a) between VE (validity of expert) and r_a, (b) between VE and the validity of all four algorithms(Logistic Regression, discriminant analysis, ID3, and neural network), (c) between r_a and the validity of all four algorithms, (d) between C index (residual correlation) and the validity of nonlinear models (ID3 and Neural Network), and (e) between VE and the incremental validity of linear models (Logistic Regression and discriminant analysis).

Correlation (a) ($r=0.95$) implied that r_a (achievement index) was a reliable indicator of the validity of the expert. Correlations (b) ($r=0.7856$, 0.7619, 0.9499, 0.8518) and (c) ($r=0.81$, 0.7789, 0.9056, 0.875) showed that model validity was significantly dependent upon expert validity. Since the models were approximations of the decisions of experts, it was not surprising that validity of models and validity of experts were closely correlated. The results suggest that the better the quality of expertise modeled, the more accurate the models become. For the ID3 algorithm, the validity of the experts was especially crucial for model performance ($r=0.9499$).

Significant correlations (d) ($r=0.5374$, 0.6271) also indicated that valid nonlinearity (measured as C) could be well modeled by the neural network and ID3. Correlation between C and validity of ID3 approached the critical value ($r=0.5374$, p-value$=0.071$). These results imply that, when valid nonlinear strategies aremodeled, it improves model performance to employ nonlinear models such as ID3 or the neural network. C index also proves to be quite useful in determining valid nonlinearity in decision behavior.

Finally, significant negative correlations (e) ($r=-0.6516$, -0.4901) between the incremental validity of linear models (LR & DA) and VE (expert's validity) are interesting. Thus, it implies that as expert's validity increases, linear models would be less likely to outperform the experts. In other words, bootstrapping using the Logistic Regression or discriminant analysis would be less possible. This result proves the early argument of Libby [16] that pointed out the limitation of linear models in expert-modeling. Bootstrapping by nonlinear algorithms may be achieved only when valid nonlinear strategy (indicated by high C index) exists dominantly in the expert's decision making behavior.

Negative correlations between the incremental validity of all models and VE showed the limitation of using a modeling algorithm, especially linear algorithms, in modeling expert decision making. When a true expert is to be modeled, the advantage of using a modeling algorithm decreases and inductive modeling should be carried out with caution. The negative correlations also showed the risk of separating the analysis

Table 4.4 A Summary of Correlation Coefficients

	(1) VE (Validity of Expert)	r_a	C
(2) r_a	0.9538* (<0.001)	1	0.5653 (0.055)
(3) C	0.5160**(0.085)	0.5653**(0.005)	1
(4) VLR	0.7856* (0.002)	0.8100* (0.001)	0.4245 (0.169)
(5) VDA	0.7619* (0.004)	0.7789* (0.002)	0.3385 (0.281)
(6) VID3	0.9499* (<0.001)	0.9056* (<0.001)	0.5374**(0.071)
(7) VNN	0.8518* (<0.001)	0.8750* (<0.001)	0.6271* (0.029)
(8) INVLR	-0.6516* (0.021)	----	----
(9) INVDA	-0.4901**(0.100)	----	----
(10) INVID3	-0.3571 (0.250)	----	----
(11) INVNN	-0.3201 (0.310)	----	----

* Critical Value : $|r|=0.58$ at 0.05 level and $N=12$.

** Critical Value : $|r|=0.50$ at 0.1 level and $N=12$.

of decision making behavior of an expert from modeling the expert. They necessitate a more comprehensive use of multiple methodologies in expert-modeling which includes process-tracing of decision behavior as well as a variety of decision analysis techniques.

4.4. Discussion

The major finding of this study was that there was a significant negative correlation between validity of experts and the incremental validity of the linear models (Logistic Regression and discriminant analysis) over the expert. The negative correlation could present a possible explanation for the conflicting results concerning bootstrapping by linear models. The negative correlation proves Libby's argument [16] that as the expert validity increased, the advantage of using linear models decreased. It implies that if the subjects were not domain experts, bootstrapping could improve the performance of models over nonexpert subjects by removing the inconsistency in human decision-making behavior.

The second finding was that significant correlations existed between the valid nonlinear index (C) and the validity of two nonlinear algorithms : the neural network ($r=0.63$, p-value<0.03) and ID3 ($r=0.54$, p-value<0.08). The significant correlations supported the use of nonlinear algorithms to capture the valid nonlinear strategy in order to improve the predictive validity of model. This finding implied that the valid nonlinear index (C) could be useful as an indicator of the limitations of linear algorithms as well as the feasibility of using nonlinear algorithms.

In comparing the two linear algorithms, LR models performed almost as well as DA models. Statistically, the difference between the two was insignificant. This result was consistent with the previous comparative study [12].

5. Conclusion

The results of this study brought an important concept, validity of decision strategy, to expert-modeling. The inclusion of the new concept, when combined with characteristics of modeling algorithms, helped explain clearly why a certain algorithm performed better than another. Therefore, the contingent relationship among validity of decision strategy, modeling algorithm, and model performance will be useful for further refinements of future expert-modeling research.

In modeling human experts, there are many factors to be considered. Previous studies using process-tracing approaches identified task characteristics, modeling situations, and individual difference as key factors [7]. In further research, the expansion of the contingent relationship to a framework which will be able to explain the relationships between model performance and those factors, may insure valuable outcomes.

REFERENCES

[1] Brunswick, E., *The Conceptual Framework of Psychology,* University of ChicagoPress, Chicago, 1952.

[2] Casey, C. J., "Prior Probability Disclosure and Loan Officers' Judgments : Some Evidence of the Impact", *Journal of Accounting Research,* 21, 1 (1983), 300-307.

[3] Chung, H. M. and M. S. Silver, "Rule-Based Expert Systems and Linear Models : An Empirical Comparison of Learning-By-Examples Methods", *Decision Sciences,* 23,3 (1992), 687-707.

[4] Dawes R. and B. Corrigan, "Linear Models in Decision Making", *Psychological Bulletin,* 81, 2 (1974), 95-106.

[5] Duliba, K., "Contrasting Neural Nets with Regression in Predicting Performance in the Transportation Industry", *Proceedings of the HICSS,* (1991), 163-170.

[6] Dutta, S. and S. Shekhar, "Bond Ratings : a Non-conservative Application of Neural Networks", *Proceedings of the ICNN,* 2 (1988), 443-450.

[7] Einhorn, H., "Expert Measurement and Mechanical Combination", *Organizational Behavior Human Performance,* 20, 2 (1972), 86-196.

[8] Einhorn, H., D. Kleimuntz, and B. Kleimuntz, "Linear Regression and Process-tracing Models of Judgment", *Psychological Review,* 86, 5 (1979), 465-485.

[9] Fisher, D. H. and K. B. McKusick, "An Empirical Comparison of ID3 and Back-propagation", Technical Report CS-88-14, Dept. of Computer Science, Vanderbilt University, 1989.

[10] Goldberg, L. R., "Man Versus Model of Man : Just How Conflicting is that Evidence ? ", *Organizational Behavior Human Performance,* 16, 1 (1976), 13-22.

[11] Hammond, K., C. Hursch, and F. Todd, "Analyzing the Components of Clinical Inference", *Psychological Review,* 71, 6 (1964), 438-456.

[12] Harrell, F. E. and K. L. Lee, "A Comparison of the Discrimination of Discriminant Analysis and Logistic Regression under Multivariate Normality", Biostatistics : in Bimedical, Public Health, and Environmental Sciences, P. K. Sen (ed.), North Holland, Amsterdam, 1985.

[13] Harris, C. R., "An Expert Decision Support System for Auditor Going Concern Evaluation", Ph.D. Dissertation, University of Texas at Arlington, 1989.

[14] Larker, D. and P. Lessig, "An Examination of the Linear and Retrospective Process Tracing Approaches to Judgment Modeling", *The Accounting Review,* 58, 1 (1983),58-89.

[15] Levi, K., "Expert Systems Should Be More Accurate Than Human Experts", *IEEE Transactions on Systems,*

Man, and Cybernetics, 19, 3 (1989), 647-657.

[16] Libby, R., "Man Versus Model of Man : Some Conflicting Evidence", *Organizational Behavior Human Performance,* 16, 1 (1976), 1-12.

[17] Libby, R., Accounting and Human Information Processing : *Theory and Applications,* Englewood Cliffs, Prentice Hall, New Jersey, 1981.

[18] Lippmann, R., "An Introduction to Computing with Neural Nets", *IEEE ASSPMagazine,* 4, 2 (1987), 4-22.

[19] Michalski, R., "A Theory and Methodology of Inductive Learning", in *MachineLearning : An Artificial Intelligence Approach,* R. Michalski, J. Carbonell, and T. Mitchell (eds.), Tioga Press, Palo Alto, 1983.

[20] Michie, D., "Current Developments in Expert Systems", in *Applications of Expert Systems,* J. Quinlan (ed.), Addison-Wesley Publishing Company, New York, 1987.

[21] Quinlan, J., "Discovering Rules by Induction from Large Collection of Examples", in *Expert Systems in Microelectronic Age,* D. Michie (ed.), Edinburgh University Press, Edinburgh, (1979), 169-201.

[22] Raghupathi, W., L. Schkade, and B. Raju, "A Neural Network Application For Bankruptcy Prediction", *Proceedings of the HICSS,* (1991), 147-155.

[23] Rumelhart, D., G. Hinton, and R. Williams, "Learning Internal Representations by Error Propagation" in *Parallel Distributed Processing,* D. Rumelhart and J. McClelland (eds.), MIT Press, Massachusetts, 1986.

[24] Shannon C., "A Mathematical Theory of Communications", *Bell Systems Technical Journal,* 27 (1948), 379-423.

[25] Tam, K. Y. and M. Kiang, "Managerial Applications of Neural Networks : The Case of Bank Failure Predictions", *Management Science,* 38 7 (1992), 926-947.

[26] Tucker, L. R., "A Suggested Alternative Formulation in the Development of Hursch, Hammond, and Hursch, and by Hammond, Hursch, and Todd", *Psychological Review,* 71, 6 (1964), 458-530.

[27] Zimmer, I., "A Lens Study of the Prediction of Corporate Failure by Bank Loan Officers", *Journal of Accounting Research,* 18, 2 (1980), 629-636.