International Journal of Computational Intelligence and Organizations
1996, Vol. 1, No. 3, 134–143

Using Neural Networks to Predict Performance of Sino–Foreign Joint Ventures

Michael Y. Hu
Department of Marketing
College of Business
Kent State University

Ming S. Hung and Murali S. Shanker
Department of Administrative Sciences
College of Business
Kent State University

Haiyang Chen
Department of Accounting and Finance
Williamson School of Business
Youngstown State University

An artificial neural network model was used to predict the performance of Sino–foreign joint ventures. Performance of international joint ventures remains a relatively underresearched area, yet its importance is well recognized due to the tremendous surge in joint-venture activities in the past decade. Data on 2,416 Sino–foreign joint ventures were gathered, allowing for empirical analysis using neural networks and traditional approaches such as logistic regression. The results showed that neural networks are indeed a viable approach to modeling the performance of joint ventures. This study identified and examined the effects of key parameters in neural networks for such modeling efforts. Moreover, in comparison to logistic regression, neural networks provided better results in classification. This article also attempts to provide a basis on which neural networks can be used for the purpose of variable selection in statistical modeling.

Key words: neural networks, classification, joint ventures

1. Introduction

Ownership of foreign operations has been a central issue in the study of multinational firms. Foreign ownership can be attained in several ways. Joint venture emerges as the dominant form of foreign investment. It allows the multinationals to lower the risks associated with investments while taking advantage of the knowledge possessed by their local partners. The primary focus in the study of international joint ventures has traditionally been on the development of theory explaining the choice of entry modes. A multinational can elect to enter a foreign market by way of exporting or forming a foreign investment enterprise. A foreign investment enterprise can also take on one of three forms—wholly owned subsidiary, equity, or contractual joint venture. A transaction cost framework typically has been used to explain these phenomena. Even though the importance of the topic has been well recognized, due to the lack of reliable data researchers have not been able to investigate fully the effects of economic factors on performance of these international joint ventures. The tremendous surge of joint-venture activities in China, particularly in the past 5 years, allows the opportunity for such investigation.

Our study attempted to explain the performance of Sino–foreign joint ventures formed between 1979 and 1990. It constitutes the first large-scale economic analysis of performance in this area. At the same time,

Michael Y. Hu, Department of Marketing, College of Business, Kent State University, Kent, OH 44242–0001. E-mail: mhu@kentvm.kent.edu.

Ming S. Hung and Murali S. Shanker, Department of Administrative Sciences, College of Business, Kent State University, Kent, OH 44242–0001. E-mail: mhung@axon.kent.edu.

Haiyang Chen, Department of Accounting and Finance, Williamson School of Business, Youngstown State University, Youngstown, OH 44555.

researchers are constantly seeking better ways to model the economic determinants of performance. In our study, we used artificial neural networks to model the relationship between performance and economic factors and empirically compared the outcome of a neural network model with one using logistic regression. Comparisons were made in terms of the ability of the models to identify significant factors and overall prediction of performance. Section 2 provides the rationale for studying the performance of Sino–foreign joint ventures and a description of the data used in this study. Section 3 introduces neural networks and specifically describes neural networks for classification problems. Results of such empirical application appear in Section 4, and conclusions come in Section 5.

2. International Joint Ventures

Formation of international joint ventures creates corporate synergism. Local partners, particularly from developing countries, would benefit from the technological know-how, management skills, and capital brought in by their foreign partners. Multinationals can benefit from their local partners' knowledge of the marketplace and position in the existing distribution network. Economies of scale can be achieved through trust and cooperation among the partners (Contractor & Lorange, 1988). Although improving efficiency through combining resources, joint ventures have also created management problems through different goals, values, and cultures. Existing literature on joint-venture performance has indicated that the success rate of international joint ventures has been "dismal, ranging from 30% to 70%" (Geringer & Hebert, 1991). Prior studies focused their attention on the performance of joint ventures in developed countries. As more joint ventures are formed in developing countries, it is important to investigate the performance of these ventures.

2.1. Sino–Foreign Joint Ventures

In the past decade, foreign investments in China have increased significantly. Previous studies have relied on interviews and surveys, securing information from foreign investors on whether they are satisfied with their operations in China (Davidson, 1987; Shenkar & Zeira, 1992; Teagarden & Von Glinow, 1990; US–China Business Council, 1990). After successfully gathering a large-sample data set from reliable sources, we sought to fill the void in the literature by conducting the first large-scale econometric study on the performance of Sino–foreign joint ventures. Multinationals extend substantial resources in the formation and implementation of these ventures. The profit potentials are huge. Multinationals have an obvious interest in accurately predicting the success of these ventures. Even a moderate improvement in model prediction will be of importance to these investors.

The literature has identified at least six major categories of factors influencing the performance of a joint venture. First, Killing (1983) suggested that the level of commitment each party brings to a joint venture is one of the key determinants of performance. A lack of commitment will lead to an ill-defined set of objectives and lack of overall direction for the organization. Beamish and Banks (1987), using a sample of 66 joint ventures in the Caribbean economies, provided some empirical evidence in support of the positive effect of commitment on performance.

Second, joint-venture performance is also determined by the amount of control exercised by the foreign partner. After reviewing the relationship between control and performance of a joint venture, Geringer and Hebert (1989) developed a conceptual framework detailing the various components of control and how each may affect performance. Ownership constitutes a key control mechanism. Beamish and Banks (1987) reported that, among 12 international joint ventures formed in less developed countries, performance is negatively related to the level of control exercised by the foreign partner.

Third, and closely related to the concept of control, is the problem of shared management. This problem is heightened when the number of partners to a joint venture increases. It is already difficult to serve under two masters, let alone more than two venture managers. Yet this problem in shared management has to be tempered with the benefits that are gained through coordination and cooperation (Gomes-Casseres, 1989). The benefits of corporate synergism outweigh the costs when the number of partners is small. Yet, as the number of partners increases, the problem of shared management may exceed the benefits derived from cooperation. Thus, a curvilinear relationship is expected between number of partners and performance.

Fourth, country of origin of foreign investment has been found to be useful in explaining performance. Killing (1983) suggested that performance is affected by the sociocultural distance between host and home countries. Killing indicated that, if a joint venture is to be successful, its managers have to develop into an effective and cohesive management team. The smaller cultural gap between the parents forming the venture, the easier it will be to create the required cohesion. Managers from similar cultures will more likely share the same values, resulting in fewer conflicts and better performance. The country-of-origin effect can also be reflected in the difference in management style adopted by managers from different countries. Child, Markoczy, and Cheung (1994) concluded that the approaches and priorities adopted by foreign managers would contrast in predictable ways, engendering different reactions from their local counterparts and setting in motion different modes of organizational learning.

Fifth, performance of international joint ventures is found to be related to the characteristics of the products involved in the ventures. Kogut (1989) studied the stability of 92 U.S.–foreign joint ventures in the United States. Stability was measured in terms of the likelihood of termination. Kogut found that high-technology products command a stronger market position in the local or foreign markets due to product differentiation. These factors allow firms to charge a higher price and engage in nonprice competition. Thus, a positive relationship is expected between the technological characteristics of a product and performance. Lecraw (1983) examined the performance of 153 joint ventures in Thailand, Malaysia, Singapore, Indonesia, and the Philippines and concluded that labor-intensive products can easily take full advantage of the abundance of low-cost labor in these developing countries. A positive relationship should persist between performance and labor intensity.

Sixth, in a developing country like China, the major business centers are better equipped to conduct business operations. In addition, the Chinese government has also established special economic and technological development and trade zones to foster the development of these activities. In these areas, the tax rates are low and facilities are plentiful. We envisioned that joint ventures should benefit from the better infrastructure and favorable tax structure. Thus, we proposed that the performance of joint ventures should vary by their business location within the different regions in China.

2.2. Data

The data for this study come from (a) the "Statement of Sino–Foreign Joint Ventures, 1979–1990" in the *Almanac of China's Foreign Economic Relations and Trade* (Ministry of Foreign Economic Relations and Trade, 1980–1991) and (b) Sino–foreign joint ventures honored for their outstanding performance by the China Association of Enterprises With Foreign Investment, a nonprofit organization for Sino–foreign joint ventures ("Commendation for Outstanding," 1991). The statement contains a description of 3,071 Sino–foreign joint ventures, providing names of the ventures, names of the parent companies, geographic location in China, number of years the venture is to exist, amount of total investment, percent of foreign investment, and primary scope of the business.

Of the 3,071 joint ventures, 2,416 fall into the manufacturing category. In 1989 and 1990, the China Association of Enterprises With Foreign Investment evaluated operating joint ventures and selected 1,187 outstanding manufacturing enterprises. To be included in the honor roll, a joint venture must (a) produce high-quality products, (b) have a before-tax profit exceeding 1 million yuans, and (c) have an export value greater than U.S. $2 million. Our study used those joint ventures commended on the honor roll as successful joint ventures because they simultaneously produce high-quality goods, profits, and export revenue. Profits understandably are important for a foreign joint venture. This measure has been adopted by various researchers (e.g., Lecraw, 1983). Producing high-quality products for exporting is also beneficial to all parties involved. For foreign partners, it is an indication that products produced in China are of good quality and competitive in international markets. Matching the 2,416 joint ventures in the statement and the 1,187 in the honor roll results in 220 of the 2,416 joint ventures receiving commendation in 1989 or 1990.

The average investment of a joint venture amounted to U.S. $3.087 million, and the average duration was 13.74 years. There was a distinct difference in the number of joint ventures formed and the average amount of foreign capital invested by the country of origin. Hong Kong investors commanded the lion's share of investment activities—60.6% of the projects amounting to 48.0% of dollar investment—and were involved in small investment projects with an average of U.S. $2.442 million. In contrast, European companies formed only 125 joint ventures; they mostly invested in larger ventures with an average of U.S. $7.675 million per venture. We noted that Japanese investors accounted for 13.7% of the number of joint ventures, amounting to roughly 18% of dollar investment; their average of U.S. $4.105 million was quite comparable to that of U.S. investors (U.S. $3.847 million).

The dependent variable, performance of joint venture, was coded as 1 for *successful* (those listed on the honor roll) and 0 otherwise. Independent variables used in this study included level of commitment, foreign control, number of partners, country of origin, industrial characteristics of joint-venture products, and joint-venture location in China. The level of commitment involved was measured using duration (number of years as specified in the joint-venture contract; DURATION) and total amount of investment. Examination of the sample data revealed that the distribution of total amount of investment contained some outlying observations. In order to minimize the effect of this source of bias, we applied the natural logarithmic transformation to this variable (LINVEST) before subjecting it to further analysis. A larger share of equity ownership would give a partner greater control of a joint venture. Percent foreign ownership (PFINVEST) was used as a proxy of the level of a foreign investor's control.

Analysis of the relationship between number of partners and performance indicated that, although the majority of joint ventures involves two partners, only 4.54% of them are successful. The success ratio increases as the number of partners increases, reflecting a positive effect of synergism. The success ratio takes a nosedive when the number of partners exceeds four. Given the curvilinear nature of the relationship, we applied the logarithmic transformation to the number of partners (LPARTNER). Five broad categories of countries were established to capture the country-of-origin effect (ORIG)—Hong Kong, Japan, United States, Europe, and other. Four dummy variables (ORIG1, ORIG2, ORIG3, ORIG4) with 1, 0, −1 coding for each were used for the first four country categories. For example, if the country being coded was Japan, then investments from this country were coded as 1; the "other" category was coded as −1; and the remaining three categories (Hong Kong, United States, Europe) were coded as 0.

In order to examine the industry-related effects, we classified the 2,416 manufacturing joint ventures into high or low technology (TECH) and labor or capital intensive (CAP). This classification scheme was developed by Dunning (1979) and later used by Lee (1983). A 1, −1 coding was used for each of the two industry-related factors (TECH, CAP). TECH was coded as 1 (−1) for high (low) technology, and CAP was coded as 1 (−1) for labor (capital) intensive. Economic risk due to differences in joint-venture location was examined at the regional level. Coastal regions offer better infrastructure and government tax incentives. Corporate income tax rate for joint venture in China is 30%. A local surtax of 10% of this income tax is levied by the local government, thus bringing the effective tax rate to a total of 33%. In the special economic zones and economic and technology development zones in the 14 cities, this tax rate is reduced to 15% for foreign investors. In this study, joint ventures formed in these zones were classified as areas with both economic and tax incentives. The coastal regions in China are more economically developed, providing facilities for conducting business operations. These were classified as regions with economic but no tax incentives. The inland provinces typically lag in economic development and were classified as areas with no tax and economic incentives. Two dummy variables (LO1, LO2) with 1, 0, −1 coding were used for the first two regions. For example, LO1 was coded as 1 for regions with both economic and tax incentives, as 0 for regions with only economic incentives, and as −1 for regions with no incentives.

Given that the dependent variable in this study is categorical in nature (either successful or not successful), the problem evolved into one of classification. That is to say, the economic factors were first used to predict whether a joint venture would belong to one or the other category, and then the resulting functional relationship was used to classify the observations into one of the two categories. Section 3 presents a brief review of neural networks and their application to classification problems.

3. Neural Networks for Classification

Neural networks have been used in business classification problems such as detecting bank failures (Tam & Kiang, 1992) and rating corporate bonds (Surkan & Singleton, 1990). Other business applications of neural networks can be found in Wilson and Sharda (1992).

3.1. Brief Review of Classification Theory

Classification is the procedure for assigning an object to a group based on observed data. In this study, each joint venture was assigned to either the successful group or the unsuccessful group. Following Hung, Hu, Shanker, and Patuwo (1996), let ω_j denote the fact that an object is a member of group j. Define

$P(\omega_j)$ = prior probability of group j, the probability that a randomly selected object belongs to group j.

$f(x \mid \omega_j)$ = conditional probability density function for x being a member of group j.

Then, using Bayes's theorem, the "posterior" probability $P(\omega_j \mid x)$, which is the probability that object x belongs to group j, can be computed as:

$$P(\omega_j \mid x) = \frac{f(x \mid \omega_j)P(\omega_j)}{\int_x f(x \mid \omega_j)P(\omega_j)}$$

The posterior probability can then be used to classify object x.

Let $c_j(x)$ be the cost of assigning x to group j. $C_j(x) = c_j(x)P(\omega_j \mid x)$ is the expected cost of assigning x to group j. The Bayesian decision rule is to assign x to the group with the least cost—namely,

$$\text{Decide } \omega_k \text{ for } x \text{ if } C_k(x) = \min_j C_j(x)$$

One example of the cost is the binary: $c_j(x) = 0$ if j is the correct group for x and 1 if it is not. Then, the decision is to assign x to the group with the highest posterior probability—namely, decide ω_k for x if $P(\omega_k \mid x) = \max_j P(\omega_j \mid x)$. This is the decision rule used in the results reported here.

Unlike the traditional methods (e.g., linear discriminant analysis, logistic regression), feedforward neural networks and other least-square methods can be used to estimate the posterior probability function directly.

3.2. Estimation of Posterior Probability With Neural Networks

The neural networks used in this study are feedforward networks with one hidden layer. The number of input nodes is equal to the number of variables in the data set. There is one output node, which contains

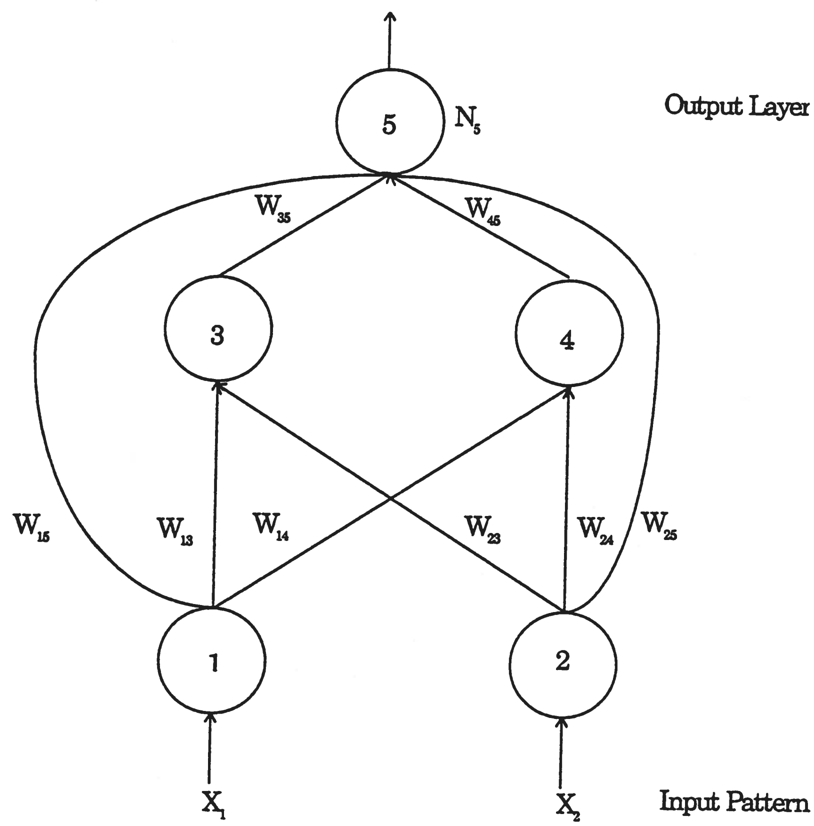

Figure 1. *Neural network with two hidden nodes. N_j = bias at node j, W_{ij} = weight on arc ij, logistic function = $(1 + e^{-x})^{-1}$.*

a bias. The activation function at the output and hidden nodes is logistic. A special feature is that there are arcs (connections) between the input nodes and the output node (Figure 1 shows an example of the type of neural network used in our study). The reason for this is to make it as close to logistic regression as possible. The logistic regression model has the form:

$$\mathrm{logit}(P(\omega_j \mid x)) = \ln \frac{P(\omega_j \mid x)}{1 - P(\omega_j \mid x)} = \beta_0 + \beta^t x$$

where β_0 is the intercept and β is the vector of regression coefficients (the superscript t denotes vector transpose). Rewriting, we get

$$P(\omega_j \mid x) = \frac{1}{1 + e^{-(\beta_0 + \beta^t x)}}$$

The right side of the equation is exactly the output of a network in Figure 1 without hidden nodes. The only difference is that, in neural network terminology, the intercept β_0 is called the *bias* and the regression coefficients in β are called the *arc weights*. Hidden nodes thus offer a convenient way to add extra modeling capabilities to the logistic regression model.

Let x_i be the vector of economic factors of joint venture i. The "target" value is $t_i = 1$ if the venture is successful and $t_i = 0$ if it is not. For a given network, let a_i be the output (there is only one output node) when x_i is the input. Training is accomplished with the following objective function:

$$\min \sum_i (a_i - t_i)^2$$

The quantity to be minimized is called the *sum of squared errors* (*SSE*). According to the theory of least-square estimation (see Hung et al., 1996, for references), the network output a_i is an unbiased estimate of the posterior probability $P(\omega_1 \mid x_i)$—namely, the probability that the venture would be successful. This theory assumes that the training sample is a "pure random" sample.

Training refers to the process of determining the weights (including biases) so that SSE is as small as possible. As this is a minimization problem, nonlinear programming methods can be used for training. Indeed, the networks are trained with the GRG2-based algorithm of Subramanian and Hung (1993). GRG2 is a widely distributed nonlinear optimization system (Lasdon & Waren, 1986). Subramanian and Hung showed that this GRG2-based algorithm is much more efficient than back-propagation.

One drawback is that, mathematically speaking, this training problem belongs to the nonconvex minimization problem for which finding the "global solution" is unsolvable. In other words, there is no known way to guarantee that it can be found. One way to increase the chances of finding a "good" solution is to use multiple starting weights.

The number of hidden nodes is an experimental factor. Theoretically, with a large number of hidden nodes, a neural network can approximate any function arbitrarily closely (see Hung et al., 1996, for an explanation). The function to be approximated here is the posterior probability function. From this point of view, more hidden nodes are desirable. On the other hand, from the point of view of statistical estimation, more hidden nodes mean fewer degrees of freedom, and the possibility of overfitting may arise. A rough analogue is multiple regression. Each time an independent variable is added to the function, 1 *df* is lost because an extra coefficient is to be estimated. For neural networks, each arc weight is an unknown to be estimated. By adding a hidden node, several extra arcs are added. Statistically speaking, a small degree of freedom can mean overfitting in that the approximation is near perfect for the sample data but is not very good for the entire population. The question of the number of hidden nodes can be answered only experimentally, with empirical data.

3.3. Model Building

The primary objective of this study was to build a model to explain joint-venture performance. As in all modeling exercises, parsimony is an important criterion. It means that one wishes to achieve the highest degree of explanatory power with the smallest model. For this objective, two related questions need to be answered:

1. What combination of variables provides the best explanation for joint-venture performance?
2. What is the appropriate neural network model to use for this data set? For this study, this means the number of hidden nodes.

The best set of variables to use for explaining joint-venture performance is determined by a backward-elimination method. Starting with the full list of variables, a variable is eliminated one at a time until some stopping criterion is met. Let a neural network model be referred to as the *full* model and its subnetwork be referred to as the *reduced* model. Here the reduced model is obtained by eliminating one input node and its associated arcs. Define

$$SSE_F = SSE \text{ of the full model}$$

$$SSE_R = SSE \text{ of the reduced model}$$

It should be that $SSE_R \geq SSE_F$ because the full model has more weights to adjust in order to reduce the objective value. In some instances, this relation might not hold due to the poor quality of solution for the full model, the reduced model, or both. To increase the chances that the relation holds, each network is trained with a large number of starting weights. The solution with the minimum *SSE* is then chosen for that particular model. This approach succeeded in producing the correct *SSE*s (as far as relative magnitude is concerned) for the models considered here.

The stopping criterion is based on the *F* ratio of regression theory:

$$F = \frac{\dfrac{(SSE_R - SSE_F)}{(df_R - df_F)}}{\dfrac{SSE_F}{df_F}}$$

where df_R and df_F refer to the degrees of freedom of the reduced and full models, respectively. Degrees of freedom, in general, are defined as number of observations minus number of weights estimated. For example, suppose there are 20 observations in the data set. If we used the network in Figure 1 in which the number of weights estimated is 9 (8 arc weights, 1 node bias), then the degrees of freedom would be 11. Clearly, $df_R \geq df_F$ because more weights are estimated in the full model compared to the reduced model.

Given a model, each input node and its associated arcs are eliminated, and the *F* ratio is calculated. The reduced model with the smallest *F* ratio is chosen. If the *F* ratio is less than a predetermined number, say F–OUT, then the corresponding input variable is eliminated. In other words, the reduced model then becomes the full model in the next iteration.

The *F* ratio has an *F* distribution only if the errors are normally distributed. There is no reason to believe that this assumption holds in neural network models. On the other hand, the *F* ratio summarizes the contribution of an input variable to the "goodness-of-fit" of a network, and so it seems like a good measure for comparing the value of variables. Section 4 presents the results of our study.

4. Results

Of the 2,416 joint ventures in the data set, approximately 70% (1,619 observations, or 147 successful + 1,472 not-so-successful joint ventures) was randomly selected for the training sample. The remainder was the *holdout* or *test sample* (797 observations, or 73 successful + 724 not-so-successful ventures).

The question of what is the appropriate network to use was answered by varying the number of hidden nodes. The values used here were from 3 to 12. Table 1 shows a subset of the results. It can be seen that, as the hidden nodes increased, the overall classification rate in the training sample increased, but it reached its

Table 1. *Classification Results for Training and Test Samples by Hidden Nodes*

	Training Sample[a]			Test Sample[b]		
	Group 1	Group 2	Overall	Group 1	Group 2	Overall
Hidden Nodes						
7	46 (31.29)	1,457 (98.98)	1,503 (92.84)	5 (6.85)	691 (95.44)	696 (87.33)
8	48 (32.65)	1,461 (99.25)	1,509 (93.21)	5 (6.85)	694 (95.86)	699 (87.70)
9	73 (49.66)	1,462 (99.32)	1,535 (94.81)	6 (8.22)	698 (96.41)	704 (88.33)
10	71 (48.30)	1,463 (99.39)	1,534 (94.72)	8 (10.96)	685 (94.61)	693 (86.95)
Logistic Regression	0 (0.00)	1,472 (100.00)	1,472 (90.92)	0 (0.00)	724 (100.00)	724 (90.84)

Note: Data are number correctly classified (percentage in parentheses).
[a]n = 1,619 (147 in Group 1, 1,472 in Group 2). [b]n = 797 (73 in Group 1, 742 in Group 2).

maximum at 9 hidden nodes in the holdout sample. The final architecture chosen had 9 hidden nodes. As shown in Table 1, this architecture produced the best classification rates for both the training and test samples.

Two questions can be raised about Table 1. First, the classification rate for Group 1 (successful ventures) is low. The best in the training sample is 49.66%; the best in the test set is only 10.96%. This issue of "small-group classification" is a common problem in statistical analysis. Indeed, the bottom of Table 1 shows that the classification rate for Group 1 by logistic regression is 0% in both the training and test samples. The explanation of the neural network results must rely on the Bayesian theory of classification in that the network output is an estimated posterior probability and the results here are simply the consequences of this estimation. In other words, the set of input variables can be used to identify 49.66% of the successful joint ventures at the best, given that the cost of misidentification–identification is equal. If one wishes to increase this classification rate, then a higher cost should be assigned. This issue is not explored here because the main objectives of this article are about model building rather than prediction.

The second question is why is there such a large difference between the training-sample and test-sample results. Take 9 hidden nodes. The classification rate for Group 1 is 49.66% in the training sample but only 8.22% in the test sample. Overall, it is 94.81% in the training sample and 88.33% in the test sample. It seems to indicate that the neural network is not a "robust" method in that its results vary widely from one application to another. To gain more insight into this question, 50 additional test samples were randomly drawn, with each sample being 50% of the full data set. For these test samples, the average number of successful joint ventures (Group 1) is 108.84, and the average number of not-so-successful ventures (Group 2) is 1,099.16 (SD = 7.08 for both groups). When these 50 new test samples were applied to the network with 9 hidden nodes, the mean results were 35.91% classification rate for Group 1, 98.37% for Group 2, and 92.74% overall (Table 2). They are all much better than the corresponding numbers in Table 1. Moreover, the standard deviation is 3.45% for the Group 1 classification rate, 0.28% for Group 2, and 0.57% overall. Thus, for these new samples, neural network (with 9 hidden nodes) yields stable and useful results. Applying these 50 test samples to logistic regression yielded 0% classification rate for Group 1, 99.87% for Group 2, and 90.87% overall (Table 2). These results are similar to those in Table 1 for logistic regression. Again, none of the successful joint ventures was correctly classified by logistic regression. A related but important lesson is that, for this kind of study, one should not rely on a small handout set to verify results of a model.

The issue of what is the appropriate set of explanatory variables to use was answered by a sequence of computations. First, backward elimination was used to

Table 2. *Classification Results Using 50 Test Data Sets*

	Test Classification Results[a]				
	Group 1		Group 2		Overall
	Number	Percentage	Number	Percentage	Percentage
Neural Network					
Average	39.06	35.91	1,081.28	98.37	92.74
SD	4.24	3.45	7.77	0.28	0.57
Logistic Regression					
Average	00.00	00.00	1,097.72	99.87	90.87
SD	0.00	0.00	6.94	0.08	0.57

[a]Nine hidden nodes.

Table 3. *Classification Results Using Neural Networks With Nine Hidden Nodes*

Variables Excluded From Model	SSE	Training Classification Percentage	Number of Input Nodes in Model	F Ratio for Variables Excluded
3A: Variables Dropped Include . . .				
None	79.42	94.81	12	
PFINVEST	91.03	93.89	11	21.78
ORIG	99.98	92.84	8	9.64
LO	93.84	93.21	10	13.52
CAP	91.81	93.39	11	23.23
TECH	92.66	93.39	11	24.83
DURATION	82.95	94.75	11	6.63
LPARTNER	90.52	94.38	11	20.82
LINVEST	94.10	93.39	11	27.53
3B: Variables Dropped Include DURATION and . . .				
PFINVEST	98.46	92.77	10	17.85
ORIG	100.37	92.34	7	7.86
LO	95.24	93.14	9	9.89
CAP	97.44	93.02	10	16.90
TECH	91.28	93.76	10	11.12
LPARTNER	90.52	94.38	10	10.41
LINVEST	105.70	92.03	10	24.64
3C: Variables Dropped Include DURATION, ORIG, and . . .				
PFINVEST	112.28	91.66	6	10.27
LO	105.94	92.53	5	7.10
CAP	106.37	92.16	6	8.42
TECH	110.97	91.79	6	9.86
LPARTNER	107.52	91.91	6	8.78
LINVEST	117.64	91.54	6	11.94
3D: Variables Dropped Include DURATION, ORIG, LO, and . . .				
PFINVEST	125.34	91.11	4	10.76
CAP	116.71	91.54	4	8.74
TECH	144.99	91.04	4	15.37
LPARTNER	117.86	91.41	4	9.01
LINVEST	146.99	90.92	4	15.84
3E: Variables Dropped Include DURATION, ORIG, LO, CAP, and . . .				
PFINVEST	127.04	91.11	3	9.92
TECH	123.02	91.35	3	9.08
LPARTNER	118.34	91.79	3	8.11
LINVEST	127.55	90.98	3	10.03
3F: Variables Dropped Include DURATION, ORIG, LO, CAP, LPARTNER, and . . .				
PFINVEST	130.75	90.92	2	9.63
TECH	146.99	90.92	2	12.67
LINVEST	146.99	90.92	2	12.67

reduce the number of variables. The network of 9 hidden nodes was chosen as the architecture, and the training sample was the data set. When all variables were included, there were 12 input nodes, and SSE_F was 79.42 (Table 3). If variable PFINVEST (percent foreign ownership) is removed, then SSE_R is 91.03. As df_F = 1,489 and df_R = 1,499, the resulting F ratio is 21.78, and the overall classification rate is 93.89%. If variable ORIG is removed, then 4 input nodes are removed simultaneously, because it is represented by four dummy variables (ORIG1, ORIG2, ORIG3, ORIG4). This results in an F ratio of 9.64. Table 3A shows that DURATION is the prime candidate for elimination because it has the lowest F ratio. Table 3B shows what happens after DURATION is removed. If, in addition, PFINVEST is removed, SSE_R becomes 98.46. Using the same original model as the full model, the F ratio for this variable is 17.85. With these two variables removed, the overall classification rate drops to 92.77%. It can be seen at this stage that ORIG is the prime candidate for elimination. Table 3C then shows the results after both DURATION and ORIG have been eliminated. If PFINVEST is also removed, then the F ratio is 10.27. This process continues until no F ratio is below

Table 4. *Classification Results Based on Final Model Chosen by Neural Network (PFINVEST, TECH, AND LINVEST)*

	Training Results			Mean Test Results Using 50 Data Sets		
Method	**Group 1**	**Group 2**	**Overall**	**Group 1**	**Group 2**	**Overall**
Neural Network	15 (10.20)	1,471 (99.93)	(91.79)	9.68 (8.90)	1,091.78 (99.33)	(91.18)
Logistic Regression	0 (0)	1,472 (100.00)	(90.92)	0 (0)	1,099.16 (100.00)	(90.99)

Note: Data are number correctly classified (percentage in parentheses).

9.00, ending up with three variables and 3 input nodes in the final, reduced model (Table 3F and Table 4).

This is an interesting result. With only three variables (and 3 of 12 input nodes), the *SSE* is nearly 50% more than the full model (with 12 variables), but the overall classification rate suffers only a small decrease. So, for the purposes of classification and explanation, this set of variables seems a reasonable choice.

To gain further insight, a linear regression model was built to establish the relationship between the network output value and the three remaining variables. For PFINVEST, the regression coefficient takes a value of $-.0005$ with a p value of .0225. This implies that joint ventures are less successful as percent foreign ownership increases. This supports the Beamish and Banks (1987) results. The average network output value for high-technology products is 0.247, as compared to 0.167 for low-technology products, suggesting a higher probability of success for high-technology products. Last, the regression coefficient for LINVEST is .035 ($p = .0001$), indicating a positive relationship between partners' financial commitment and performance of Sino–foreign joint ventures. The coefficients for logistic regression for these factors are $-.0119$ ($p = .1007$) for PFINVEST, $-.2628$ ($p = .0078$) for TECH, and .3948 ($p = .0001$) for LINVEST. Note that logistic regression provides the same sign as neural networks for PFINVEST and LINVEST, whereas the sign for TECH is contrary to what has been reported in previous studies.

The predictive power of the three-variable model recommended by neural network was evaluated with the 50 random samples used for Table 2. As shown in Table 4, the overall classification rate is 91.79% in the training sample and a mean of 91.18% in the test samples. The classification rate for Group 1 is only 8.90% (Table 4), compared to 35.91% in Table 2. This suggests that the removed variables are useful for identification of successful joint ventures. So, if the interest is in explaining successful joint ventures only, then all 12 variables should be retained.

To check the final model recommended by neural networks, a similar backward-elimination method was applied in the logistic-regression models. The variable with the largest p value was removed, and the regression function was recomputed. This procedure was repeated until no p value is greater than .05. The results, as shown in Table 5, have four variables: ORIG, TECH, LPARTNER, and LINVEST. With this set of variables, a neural network (9 hidden nodes) was trained and tested. It shows that, overall, neural network results are comparable to logistic regression, with slightly better results for Group 1 classification.

These results indicate that both neural networks and logistic regression are comparable. But logistic regression fails to identify any successful joint ventures in any model. But the more important thing is that neural networks produce estimates of posterior probabilities directly. With those values, one can vary the cost of misclassification–classification to achieve desirable classification rates.

5. Conclusions

The use of neural networks in business applications is just now emerging. In this study, three novel approaches were developed:

1. A procedure, backward elimination, for building a parsimonious neural-network model.
2. Use of repeated, independent test samples.
3. Use of neural-network output for regression analysis to measure the effect of an independent variable.

Regarding the application to joint ventures, the results indicate that, although successes and failures can be predicted with high accuracy, the prediction for the successful joint venture is not as good. There can be many reasons behind this phenomenon. The first is

Table 5. *Classification Results Based on Final Model Chosen by Logistic Regression (ORIG, TECH, LPARTNER, LINVEST)*

	Training Results			Test Results		
Method	**Group 1**	**Group 2**	**Overall**	**Group 1**	**Group 2**	**Overall**
Neural Network	29 (19.73)	1,463 (99.39)	(92.16)	5 (6.85)	711 (98.20)	(89.84)
Logistic Regression	1 (0.68)	1,470 (99.86)	(90.86)	0 (0)	723 (99.86)	(90.72)

Note: Data are number correctly classified (percentage in parentheses).

the definition of *success*. In this data set, *success* is defined by the government and thus reflects the concern of the Chinese government. The second reason is data availability and integrity. Some of the explanatory variables may not be included because they were not available. In addition, the reported values may contain some inaccuracies. But these two difficulties are common to all modeling approaches.

The neural-network model suggests that joint-venture performance is inversely related to percent foreign ownership, that high-technology products are more likely to be successful, and that the performance of a joint venture goes up as the level of financial commitment increases. These results agree with those of previous studies.

References

Beamish, P., & Banks, J. (1987, Summer). Equity joint ventures and the theory of the multinational enterprise. *Journal of International Business Studies*, pp. 1–16.

Child, J., Markoczy, L., & Cheung, T. (1994). Managerial adaptation in Chinese and Hungarian strategic alliances with culturally distinct foreign partners. In S. Stewart (Ed.), *Advances in Chinese industrial studies* (Vol. 4, pp. 211–232). Greenwich, CT: JAI.

Commendation for outstanding foreign investment enterprises in China from China Association of Enterprises With Foreign Investment. (1991, July 16). *People's Daily.*

Contractor, F., & Lorange, P. (1988). *Cooperative strategies in international business*. Boston: Lexington.

Davidson, W. (1987). Creating and managing joint ventures in China. *California Management Review, 29*, 77–94.

Dunning, J. (1979, November). Explaining changing patterns of international production: In defence of the eclectic theory. *Oxford Bulletin of Economics and Statistics*, pp. 269–296.

Geringer, M., & Hebert, L. (1989). Control and performance of international joint ventures. *Journal of International Business Studies, 20*, 233–253.

Geringer, M., & Hebert, L. (1991). Measuring performance of international joint ventures. *Journal of International Business Studies, 22*, 249–263.

Gomes-Casseres, B. (1989). Ownership structures of foreign subsidiaries, theory and evidence. *Journal of Economic Behavior and Organization, 11*, 1–25.

Hung, M. S., Hu, M. Y., Shanker, M. S., & Patuwo, B. E. (1996). Estimating posterior probabilities in classification problems with neural networks. *International Journal of Computational Intelligence and Organizations, 1*, 49–60.

Killing, P. (1983). *Strategies for joint venture success*. New York: Praeger.

Kogut, B. (1989). The stability of joint ventures: Reciprocity and competitive rivalry. *Journal of Industrial Economics, 38*, 183–198.

Lasdon, L. S., & Waren, A. D. (1986). *GRG2 user's guide*. Austin: University of Texas, School of Business Administration.

Lecraw, D. (1983). Performance of transnational corporations in less developed countries. *Journal of International Business Studies, 14*, 15–33.

Lee, C. (1983). International production of the United States and Japan in Korean manufacturing industries: A comparative study. *Weltwirtschaftliches Archiv, 119*, 745–753.

Ministry of Foreign Economic Relations and Trade. (1980–1991). Statement of Sino–foreign joint ventures, 1979–1990. In *Almanac of China's foreign economic relations and trade*. Beijing: Author.

Shenkar, O., & Zeira, Y. (1992). Role conflict and role ambiguity of chief executive officers in international joint ventures. *Journal of International Business Studies, 23*, 55–75.

Subramanian, V., & Hung, M. S. (1993). A GRG2-based system for training neural networks: Design and computational experience. *ORSA Journal on Computing, 5*, 386–394.

Surkan, A., & Singleton, J. (1990). Neural networks for bond rating improved by multiple hidden layers. *International Joint Conference on Neural Networks, 2*, 157–162.

Tam, K. Y., & Kiang, M. Y. (1992). Managerial applications of neural networks: The case of bank failure predictions. *Management Science, 38*, 926–947.

Teagarden, M., & Von Glinow, M. (1990). Sino–foreign strategic alliance type and related operating characteristics. *International Studies of Management and Organization, 20*, 99–108.

US–China Business Council. (1990). *Special report on U.S. investment in China*. Washington, DC: China Business Forum.

Wilson, R. L., & Sharda, R. (1992). Neural networks. *OR/MS Today, 19*(4), 36–42.

Author Notes

Michael Y. Hu is a Professor in the Department of Marketing, Kent State University. His recent writings have been on the topics of joint ventures, neural networks, and frequent-flier programs.

Ming S. Hung is Professor Emeritus in the Department of Administrative Sciences, Kent State University. His research interests include optimization, neural networks, and decision modeling.

Murali S. Shanker is an Assistant Professor in the Department of Administrative Sciences, Kent State University. He received a PhD from the Department of Operations and Management Science, University of Minnesota. His research interests include parallel simulation and neural networks.

Haiyang Chen is a Professor in the Department of Finance, Williamson College of Business Administration, Youngstown State University. His numerous articles on Sino–foreign joint ventures have appeared in *Journal of International Business Studies, Journal of Business Research, Financial Management*, and other publications.