Using Neural Networks To Predict The Performance of Sino-Foreign Joint Ventures

Michael Y. Hu
Department of Marketing, College of Business
Kent State University, Kent, OH 44242-0001
Phone: (216) 672-2750 Ext. 326
Fax: (216) 672-2448
E-mail: mhu@kentvm.kent.edu

Ming S. Hung
Department of ADMS, College of Business
Kent State University, Kent, OH 44242-0001

Murali S. Shanker
Department of ADMS, College of Business
Kent State University, Kent, OH 44242-0001

Haiyang Chen
Youngstown State University
Youngstown, OH

October 22, 2000

Abbreviated Title: Predicting joint ventures with neural networks

Abstract

An arti…cial neural network model is used to predict the performance of Sino-foreign joint ventures. Performance of international joint ventures remains a relatively under-researched 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 was gathered, allowing for empirical analysis using neural networks and traditional approaches such as logistic regression. The results show that neural networks are indeed a viable approach to modeling the performance of joint ventures. This study identi…es and examines the e¤ects of key parameters in neural networks for such modeling e¤orts. Moreover, in comparison to logistic regression, neural networks provided better results in classi…cation. This paper also attempts to provide a basis upon which neural networks can be used for the purpose of variable selection in statistical modeling.

Keywords: Neural networks, classi…cation, 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 a number of 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 five years, allows the opportunity for such investigation.

This study attempts 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, researchers are constantly seeking better ways to model the economic determinants of performance. In this paper, we employ artificial neural networks to model the relationship between performance and economic factors and empirically compare the outcome of a neural network model with one using logistic regression. Comparisons will be made in terms of their ability to identify significant factors and overall prediction of performance. The next section 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 are in section 4, and conclusions are 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 capitals 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 and Lorange, 1988). While improving efficiency through combining resources, joint ventures have also created management problems through different goals, values and cultures. Existing literature on joint venture performance indicates the success rate of international joint ventures has been "dismal, ranging from 30% to 70%" (Geringer and 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 rely on interviews and surveys, securing information from foreign investors on whether or not they are satisfied with their operations in China (Davidson, 1987; Shenkar and Zeira, 1992; Teagarden and Von Glinow, 1990; US-China Business Council, 1990). After successfully gathering a large-sample data set from reliable sources, we 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. They 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. Killing (1983) suggests that the level of commitment each party brings to a joint venture to be 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) based on a sample of 66 joint ventures in the Caribbean economies provide some empirical evidence in support of the positive effect of commitment on performance.

Secondly, 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) develop a conceptual framework detailing the various components of control and how each may in term affect performance. Ownership constitutes a key control mechanism. Beamish and Banks (1987) report that among twelve international joint ventures formed in less developed countries, performance is negatively related to the level of control exercised by the foreign partner.

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 when more than two parents are managing the venture. 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.

Country of origin of foreign investment has been found to be useful in explaining performance. Killing (1983) suggests that performance is affected by the socio-cultural distance between the host and home countries. He indicates 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) conclude 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.

Performance of international joint ventures is found to be related to the characteristics of the products involved in the ventures. Kogut (1989) studies the stability of 92 US-foreign joint ventures in the United States. Stability is measured in terms of the likelihood of termination. It is found that high technology products due to product differentiation command a stronger market position in the local or foreign markets. These factors allow firms to charge a higher price and engage in nonprice competition. Thus a positive relationship is expected between technological characteristics of a product and performance. Lecraw (1983) examines the performance of 153 joint ventures in Thailand, Malaysia, Singapore, Indonesia and the Philippines and concludes 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.

Lastly, 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. It is envisioned that joint ventures should benefit from the better infrastructure and favorable tax structure. Thus it is proposed that the performance of joint ventures should vary by their business location within the different regions in China.

2.2 The Data

The data for this study come from two sources: (1) Statement of Sino-Foreign Joint Ventures, 1979-1990 in Almanac of China's Foreign Economic Relations and Trade (Ministry of Foreign Economic Relations and Trade) and (2) 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 (People's Daily, 1991). The Statement contains a description of 3,071 Sino-foreign joint ventures, providing names of the ventures, their parent companies' names, geographic location in China, number of years the venture is to exist, amount of total investment, percent of foreign investment, and the primary scope of the business.

Of the 3,071 joint ventures, 2,416 fall in 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 (1) produce high quality products; (2) have a before tax profit exceeding one million yuans; and (3) export value greater than two million US dollars. This study uses those commended on the Honor Roll as successful joint ventures because they must simultaneously produce high quality goods, profits, and export revenue. Profits understandably are important for a foreign joint venture. This measure was adopted by various researchers (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 amounts to US$3.087 million and the average duration is 13.74 years. There is a distinct difference in the number of joint ventures formed and the average amount invested by the country of origin of foreign capital. Hong Kong investors command the lion's share of investment activities, 60.6% of the projects amounting to 48.0% of dollar investment. They are involved in small investment projects with an average of US$2.442 million. In contrast, European companies formed only 125 joint ventures. They mostly invested in larger ventures with an average of US$7.675 million per venture. It is noted that Japanese investors accounted for 13.7% of the number of joint ventures, amounting to roughly 18% of dollar investment. Their average of US$4.105 million is quite comparable to that of US investors, being US$3.847 million.

The dependent variable, performance of joint venture, is coded as 1 for "successful", those listed on the Honor Roll, and 0 otherwise. Independent variables used in this study include 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 is measured using duration (number of years as specified in the joint venture contract, DURATION), and total amount of investment. An examination of the sample data reveals that the distribution of total amount of investment contains some outlying observations. In order to minimize the effect of this source of bias, the natural logarithmic transformation is applied 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) is used as a proxy of the level of a foreign investor's control.

An analysis of the relationship between the number of partners and performance indicates that while the majority of joint ventures involves two partners, only 4.54% of them are successful. The success ratio increases as the number of partners goes up, reflecting a positive effect of synergism. The success ratio takes a nose-dive when the number of partners exceed four. Given the curvilinear nature of the relationship, the logarithmic transformation is applied to the number of partners (LPARTNER). Five broad categories of countries are established to capture the country of origin effect (ORIG) — Hong Kong, Japan, US, Europe and other. Four dummy variables (ORIG1, ORIG2, ORIG3, ORIG4) with 1, 0, -1 coding for each are used for the first four country categories. For example, if the country being coded is Japan, then investments from this country will be coded as "1", the "other" category as "-1" and the remaining three categories (Hong Kong, US and Europe) as "0".

In order to examine the industry related effects, the 2,416 manufacturing joint ventures are classified 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 is used for each of the two industry related factors (TECH, CAP). TECH is coded as "1" ("-1") for high (low) technology, and CAP as "1" ("-1") for labor (capital) intensive. Economic risk due to differences in joint venture location is 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 the 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 are 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 are classified as region with economic but no tax incentives. The inland provinces typically lag in economic development and are classified as areas with no tax and economic incentives. Two dummy variables (LO1, LO2) with 1, 0, -1 coding are used for the first two regions. For example, LO1 is 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 evolves into one of classification. That is to say, the economic factors will be first used to predict whether a joint venture belongs to one or the other category, and then the resulting functional relationship will be used to classify the observations into one of the two categories. The next section presents a brief review on neural networks and its application to classification problems.

3 Neural Networks For Classification

Neural networks have been used in business classification problems such as detecting bank failures (Tam and Kiang, 1992) and rating of corporate bonds (Surkan and 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 is to be assigned to either the successful group or the unsuccessful one. Following Hung, et al. (1996), let ω_j denote the fact that an object is a member of group j. Define

$P(\omega_j)$ = the prior 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 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,

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

One example of the cost is the binary: $c_j(x)$ is 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 $P(\omega_k|x) = \max_j P(\omega_j|x)$. This is the decision rule used in the results reported here.

Unlike the traditional methods such as linear discriminant analysis or logistic regression, feed-forward 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 employed in this study are feed-forward 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 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:

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

where β_0 is the intercept and β is the vector of regression coefficients. The superscript t denotes vector transpose. Rewriting the above, we have

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

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

(Figure 1 about here)

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 $P(\omega_1|x_i)$; 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. Since 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 and Waren, 1986). Subramanian and Hung (1993) show that this GRG2-based algorithm is much more efficient than back propagation.

One drawback is that mathematically speaking this training problem belongs to the non-convex 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, one degree of freedom 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 only be answered experimentally with empirical data.

3.3 Model Building

The primary objective of this study is 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:

- What combination of variables will provide the best explanation for joint venture performance?
- 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 sub-network 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 of the full model

SSE_R = SSE of the reduced model

It should be that $SSE_R \geq SSE_F$ since the full model has more weights to adjust in order to reduce the objective value. In some instances, this relation might not hold because of the poor quality of solution for either the full or 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, which is

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

where df_R and df_F refer to the degrees of freedom of the reduced and full models, respectively. The degrees of freedom, in general, is defined as df = (number of observations - 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 and one node bias), then the degrees of freedom would be 11. Clearly, $df_R \geq df_F$; since 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. The next section presents the results of our study.

4 Results

A training sample was randomly selected from the data set. Recall that there are 2,416 joint ventures in the data set, approximately 70% of them formed the training sample, giving a total of 1,619 observations with 147 successful and 1,472 not-so-successful joint ventures in the training sample. The remainder is called the holdout or test sample, which has 797 observations: 73 successful and 724 not-so-successful ventures.

The question of what is the appropriate network to use is 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 increase, the overall classi…cation rate in the training sample increases, but it reaches its maximum at 9 hidden nodes in the holdout sample. The …nal architecture chosen has 9 hidden nodes. As shown in Table 1, this architecture produces the best classi…cation rates for both the training and test samples.

(Table 1 about here)

Two questions can be raised about Table 1. First, the classi…cation rate for Group 1 (the successful ventures) is low. The best in the training sample is 49.66%, and in the test set, only 10.96%. This issue of "small group classi…cation" is a common problem in statistical analysis. Indeed, the bottom of Table 1 shows that the classi…cation rate for Group 1 by logistic regression is 0% in both the training and the test samples. The explanation of the neural network results must rely on the Bayesian theory of classi…cation 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 mis-identi…cation/identi…cation is equal. If one wishes to increase this classi…cation rate, then a higher cost should be assigned. This issue is not explored here because the main objectives of this paper are about model building rather than prediction.

The second question is "why is there such a large di¤erence between the training sample and the test set results?" Take 9 hidden nodes. The classi…cation 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. It seems to indicate that 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, additional 50 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 1099.16 for the other group, with a standard deviation of 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 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 that in Table 1 for logistic regression. Again, none of the successful joint ventures were 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.

(Table 2 about here)

The issue of what is the appropriate set of explanatory variables to use is answered by a sequence of computations. First, backward elimination was used to reduce the number of variables. The network of 9 hidden nodes is chosen as the architecture and the training sample is the data set. When all variables are included, there are 12 input nodes, and the sum of squared errors SSE_F is 79.42 (Table 3). If variable PFINVEST (percent foreign ownership) is removed, then SSE_R is 91.03. The degrees of freedom are $df_F = 1; 489$ and $df_R = 1; 499$. This results in an F-ratio of 21.78 and an overall classification rate of 93.89%. If variable ORIG is removed, then four input nodes are removed simultaneously, because it is represented by four dummy variables ORIG1, ORIG2, ORIG3 and 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, the 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, 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 9.00, ending up with three variables and three input nodes in the final reduced model (Table 3F and 4).

(Table 3 about here)

This is an interesting result. With only three variables (and 3 out of 12 input nodes), the sum of squared errors 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 is built to establish the relationship between the network output value and the three remaining variables. For PFINVEST, the regression coefficient takes a value of -0.0005 with a p-value of 0.0225. This implies that joint ventures are less successful as percent foreign ownership increases. This supports 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. Lastly, the regression coefficient for LINVEST is 0.035 (p-value = 0.0001), indicating a positive relationship between partners' financial commitment and performance of Sino-foreign joint ventures. The coefficients for logistic regression for the above factors are -0.0119 (p-value = .1007) for PFINVEST, -0.2628 (p-value = 0.0078) for TECH and 0.3948 (p-value = 0.0001) for LINVEST. Note that logistic regression provides the same sign as neural networks for PFINVEST and LINVEST, while 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 is 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.

(Table 4 about here)

To check the final model recommended by neural networks, a similar backward elimination method is applied in the logistic regression models. The variable with the largest p-value is removed and the regression function recomputed. This procedure is repeated until no p-value is greater than 0.05. The results, as shown in Table 5, has 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.

(Table 5 about here)

The above 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, several novel approaches were developed:

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

Regarding the application to joint ventures, the results indicate that while 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 one is 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 agree with the results of previous studies.

References

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

Child, J., L. Markoczy & T. Cheung(1994). Managerial adaptation in Chinese and Hungarian strategic alliances with culturally distinct foreign partners. Advances in Chinese Industrial Studies, edited by Sally Stewart, 4, 211-232 (JAI press).

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

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

Dunning, J (1979). Explaining changing patterns of international production: In defence of the eclectic theory. Oxford Bulletin of Economics and Statistics, November, 269-96.

Geringer, M. & L. Hebert (1989). Control and performance of international joint ventures. Journal of International Business Studies, 20, 233-53.

Geringer, M. & L. Hebert (1991) . Measuring performance of international joint ventures. Journal of International Business Studies, 22, 249-63.

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

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

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

Kogut, B (1983). The stability of joint ventures: Reciprocity and competitive rivalry. The Journal of Industrial Economics, 38, 183-98.

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

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, 745-53.

Ministry of Foreign Economic Relation and Trade, various editions. Almanac of China's foreign economic relations and trade, Beijing, China.

People's Daily (1991). Commendation for outstanding foreign investment enterprises in China from China Association of enterprises with foreign investment, July 16.

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

Subramanian V. & M. S. Hung (1993). A GRG2-based system for training neural networks: Design and computational experience. ORSA Journal on Computing, 5 (4), 386–394.

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

Tam, K.Y. & M.Y. Kiang (1992). Managerial applications of neural networks: The case of bank failure predictions. Management Science, 38(7), 926-947.

Teagarden, M. & M. Von Glinow (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. The China Business Forum, Washington D.C.

Wilson, R.L. & R. Sharda (1992). Neural networks. OR/MS Today, August, 36-42.

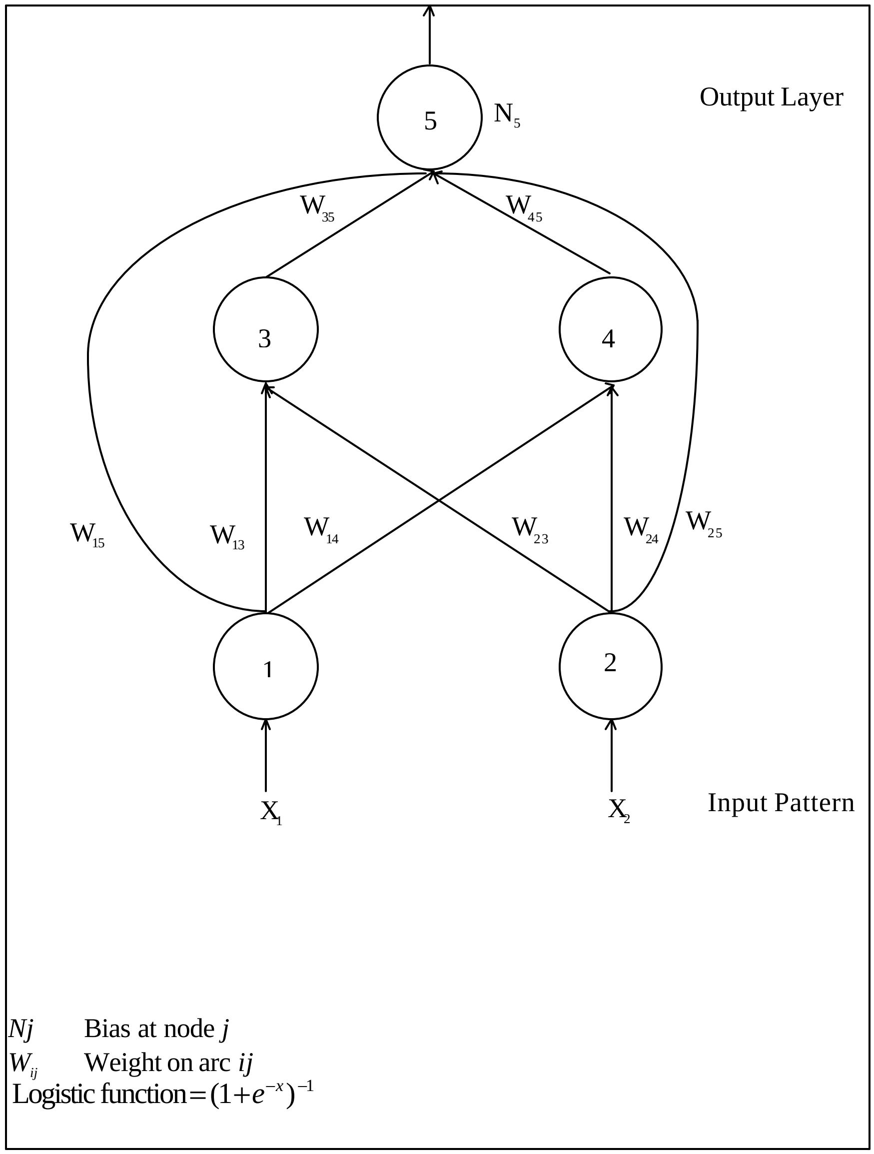

Figure 1: A Neural Network with 2 Hidden Nodes

Hidden Nodes	Training[1] Group 1	Training[1] Group 2	Training[1] Overall	Test[1] Group 1	Test[1] Group 2	Test[1] Overall
7	46[2] (31.29)	1457 (98.98)	1503 (92.84)	5 (6.85)	691 (95.44)	696 (87.33)
8	48 (32.65)	1461 (99.25)	1509 (93.21)	5 (6.85)	694 (95.86)	699 (87.70)
9	73 (49.66)	1462 (99.32)	1535 (94.81)	6 (8.22)	698 (96.41)	704 (88.33)
10	71 (48.30)	1463 (99.39)	1534 (94.72)	8 (10.96)	685 (94.61)	693 (86.95)
Logistic Regression	0 (0.00)	1472 (100.00)	1472 (90.92)	0 (0.00)	724 (100.00)	724 (90.84)

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

1 Training: 1619 (147 in Group 1, 1,472 in Group 2)
Test: 797 (73 in Group 1, 724 in Group 2)

2 Number correctly classified; percentage in parentheses

	Neural Networks (9 Hidden Nodes) Test Classi…cation Results				
	Group 1		Group 2		Overall
	Number	Percentage	Number	Percentage	Percentage
Average	39.06	35.91	1081.28	98.37	92.74
Standard Deviation	4.24	3.45	7.77	0.28	0.57

	Logistic Regression Test Classi…cation Results				
	Group 1		Group 2		Overall
	Number	Percentage	Number	Percentage	Percentage
Average	00.00	00.00	1097.72	99.87	90.87
Standard Deviation	0.00	0.00	6.94	0.08	0.57

Table 2: Classi…cation Results using 50 Test Data Sets

Variables Excluded From the Model	Sum of Squared Error	Training Classi...cation Percentage	# of Input Nodes in the 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

Table 3: Classi...cation Results Using Neural Networks with 9 Hidden Nodes

Method	Training Results			Mean Test Results Using 50 Data Sets		
	Group 1	Group 2	Overall	Group 1	Group 2	Overall
Neural Network	15 (10.20)[¤¤]	1471 (99.93)	(91.79)	9.68 (8.90)	1091.78 (99.33)	(91.18)
Logistic Regression	0 (0)	1472 (100.00)	(90.92)	0 (0)	1099.16 (100.00)	(90.99)

Table 4: Classi...cation results based on ...nal model chosen by neural network (PFINVEST, TECH, and LINVEST)

** Number correctly classi...ed; percentage in parentheses

	Training Results			Test Results		
Method	Group 1	Group 2	Overall	Group 1	Group 2	Overall
Neural Network	29 (19.73)[¤¤]	1463 (99.39)	(92.16)	5 (6.85)	711 (98.20)	(89.84)
Logistic Regression	1 (0.68)	1470 (99.86)	(90.86)	0 (0)	723 (99.86)	(90.72)

Table 5: Classi...cation results based on ...nal model chosen by logistic regression (ORIG, TECH, LPARTNER and LINVEST)

** Number correctly classi...ed; percentage in parentheses