7-1-2001

Using Neural Networks to Discover Patterns in International Equity Markets: A Case Study

Mary Malliaris
Loyola University Chicago

Linda Salchenberger
Marquette University, linda.salchenberger@marquette.edu

Published version. "Using Neural Networks to Discover Patterns in International Equity Markets: A Case Study," in *Neural Networks in Business: Techniques and Applications.* Eds. Kate A. Smith, Jatinder N.D. Gupta. Hershey, PA: IGI Global, 2001: 205-219. DOI. © 2001, IGI Global, www.igi-global.com. Posted by permission of the publisher.
Linda Salchenberger was affiliated with Loyola University Chicago at the time of publication.

Chapter XIII

Using Neural Networks to Discover Patterns in International Equity Markets: A Case Study

Mary E. Malliaris and Linda Salchenberger
Loyola University Chicago, USA

BACKGROUND

In the current economic environment, international stock markets have become increasingly linked, due to financial deregulation, the globalization of markets, and information technology. Financial deregulation and globalization of markets have contributed to the stronger relationships between stock markets, with the U.S. causing other market movements by influencing key underlying macroeconomic variables that cause stock index movements (Nasseh & Strauss, 2000). Information technology accelerates responsiveness to world events as information travels around the world in nanoseconds. Traders can now react instantly to corporate announcements, rumors, and the activities of other markets. In fact, in addition to extending trading hours to create linkages between international markets, exchange links have been developed that allow traders to trade at another exchange through specific agreements during any time of the day or night (Cavaletti, 1996).

An understanding of the international market equity structure has become important for investment decision-making, since international diversification is a strategy often used in portfolio management to reduce risk (Malliaris, 1996; Theodossiou, 1994). International portfolio diversification has gained popularity as an investment strategy in industrial countries as individual and corporate investors have been encouraged to increase their holdings in foreign securities. Barron's reported that U.S. investors had tripled their ownership of foreign equities from $63 billion in 1988 to $200 billion in 1993, and this number has more than tripled again since then. There is a widespread belief that profit can be increased and risk can be reduced with a portfolio consisting of domestic and foreign investments. The reason that international diversification was originally recommended as an investment strategy is that, if domestic and foreign markets are highly uncorrelated, portfolio risk is reduced. If, on the other hand, domestic market movement follows movement in foreign markets, the trader has some advance warning of profitable positions to take daily.

Recent events in the stock market have dramatically demonstrated the degree of integration among international equity market price indices in times of great financial upheaval. For example, the U.S. equity markets responded to the October 1997 collapse of the Southeast Asian markets with its own downward plunge, followed by the current period of volatility, demonstrating global linkages between these two markets (Lee & Kim, 1994). While it is interesting to study the periods immediately before and after such catastrophic events, the more general question is whether or not international equity markets demonstrate co-movement on a daily basis. The degree to which markets are integrated or segmented internationally and move together on a daily basis is one that impacts investment decisions for investors and traders and yet largely remains an unanswered question. Studies have suggested (Dickinson, 2000; Masih & Masih, 1997) that the U.S. has a greater impact on the other international equity markets. But the amount of day- to-day impact of the other markets on the U.S. has yet to be demonstrated. This is the problem we want to address.

The use of neural networks represents a new approach to how this type of problem can be investigated. The economics and finance literature is full of studies that require the researcher to prespecify the exact nature of the relationship and select specific variables to test. In this study, we use a multistage approach that requires no prespecification of the model and allows us to look for associations and relationships that may not have been considered. Previous studies have been limited by the nature of statistical tools, which require the researcher to determine the variables, time frame, and markets to test. An intelligent guess may lead to the desired outcome, but neural networks are used to produce a more thorough analysis of the data, thus improving the researcher's ability to uncover unanticipated relationships and associations.

PURPOSE OF THE CASE STUDY

The purpose of this case study is to step through the process of using data mining and neural networks to look for the influence of selected Eastern markets (Japan, Hong Kong, and Australia) on the S&P 500. The neural network results will be compared to a standard benchmark, the random walk forecast. The random walk hypothesis assumes that, since tomorrow cannot be predicted, the best guess we can make is that tomorrow's price will be the same as today's price. If neural networks outperform the random walk, then it can be concluded that there is a nonlinear function or process inherent in the data tested. The implication is that a short-term forecast can be successfully generated.

If the neural network/random walk contest is decided in favor of the neural network, this might lead to the conclusion that the benefits from diversification have been largely overstated and may be due to differences in real growth rates, inflation, and exchange rates. In addition, the degree to which international markets are linked may be significant for determining the cost of capital for international projects and the formulation of national economic policies. If the random walk cannot be beaten, then these results would support those investors who believe that international diversification offers protection for their portfolio.

This application is appropriate for neural networks because other statistical methods have failed to yield a good short-term prediction model. While there is some intuition regarding which market indicators may influence the price we are trying to predict, empirical validation is the ultimate test of economic theories.

MEASURING NETWORK PERFORMANCE

A variety of performance methods are used for determining the reliability, validity, and usefulness of neural networks developed for financial applications. Since the purpose of this case study is to focus on short-term forecasting, we measured the overall prediction error and compared the performance of the neural networks with the random walk, which served as a convenient bench- mark. The success of short-term trading strategies in international markets typically rely on prediction accuracy, in which unit errors are more important than percentage errors, so we have computed mean absolute error (MAE) and root mean squared error (RMSE) to measure prediction performance. The objective was to evaluate consistency and accuracy over a short time, rather than the ability to predict long-term trends and major price shifts. Other measures typically used include mean absolute percentage error and directional symmetry. For a more detailed discussion of each of these and other perfor- mance metrics, see Azoff (1994) and chapter one of this book.

METHODOLOGY

Database

A database of international market prices from 1997-2000 (open, high, low, and closing) was developed for this research and included prices for the S&P 500, Nikkei 225, Hong Kong Hang Seng, and Australian All Ordinaries. Individual data sets for each market were downloaded from Yahoo (www.yahoo.com/m2) and combined into one data set by matching on date. For each market, the following variables were available from each market for each day of the three year period: date, high price, low price, open price, and closing price. Since all markets are not open on the same set of days, only days that all markets were open were used.

Preliminary correlation analysis of the data led to the decision to focus on predicting the S&P 500 of the U.S., using the Australian All Ordinaries (AORD), the Hong Kong Hang Seng (HIS), and the Nikkei 225 of Japan (N225). The hypothesis under investigation for this case study is the effect of the three non-Western markets on the U.S. price within one day. In particular, the relationship of the difference between the open and closing price and the high and low prices in these markets was studied. This difference is used to indicate the amount of market change for a particular day in that market, and input from these markets are used to discover the impact of their changes on the U.S. market.

The impact on the open price in the U.S. can often be significantly greater or smaller if the markets are closed for any period of time. Therefore, an important variable developed and included as part of the data set is the number of nights each market was closed locally. For example, if a market is open Monday and Tuesday, the Tuesday value for this variable would be 1. In addition, correlation analysis showed that a one-day lag of Australia's price information was more highly correlated with the S&P 500 price than the current Australian price. Thus, the one-day lag of the Australian price variables was included in all data sets used.

Preparing the Data Set

The process of cleaning and preprocessing the data for neural networks is an essential first step for the development, training, and testing of the neural networks. Data preparation involves handling missing data, proper coding of data, identifying outliers, and discarding erroneous data. This data set was examined for missing values and inconsistent data or outliers. Outliers cause problems in financial prediction problems because the network cannot predict their behavior, since there are typically an insufficient number of them. Problem domain knowledge, data plots, computations of statistical measures, such as the mean and standard deviation, and histograms were used help to identify outliers. These were removed from the data set by deleting the entire row for that day. Values that were outside of a 3% range for market prices were discarded. The final data set contained 657 observations for years 1997-1999 and 239 observations for 2000.

Preprocessing the Data Set

Preprocessing the data set reduces noise and enhances the signal, thus improving the learning capability of a neural network. In many cases, combining input values reduces the input space and improves the mapping of the model. Better results are always achieved when noise is eliminated from the data set by reducing the input space and identifying and selecting variables that have the greatest impact on the output variable. Raw price data (open price, closing price, high, low) is seldom effective in neural networks, because the values overwhelm the network and it is difficult for the network to learn the trends and subtle price movements. Thus, the first step in preprocessing this data set was to convert the price data into meaningful ratios. For this data set, we used percentage of change in a market's price by computing the ratio of the difference between open and closing price divided by the open price, and similar ratios were computed for the high and low prices.

The final variable set included each of the following, computed daily for the Australian, Hong Kong, and Japanese markets:

(open price – closing price) / open price
an additional lag of (open price – closing price) / open price for Australia
(high – low) / open price
number of nights a market was closed locally
number of nights S&P 500 was closed locally

to predict

(open price – closing price) / open price for the S&P 500 on the same day.

Thus, we are looking for the sensitivity of today's S&P 500 change to changes that have happened a few hours earlier in other markets.

The next step in preprocessing the data is to normalize or scale the data. We used a technique called min-max normalization performed by the neural networks software package. This normalization process performs a linear transformation on the original input set into a specified data range, in this case 0 to 1.

Selecting a Neural Network Model

Model specification is a key issue in developing any approximation technique, and selecting a particular neural network architecture corresponds to making assumptions about the space of the approximating function. Backpropagation neural networks were selected for this problem domain because of past success with this approach for prediction problems (Malliaris & Salchenberger, 1993; Malliaris & Salchenberger, 1996a; Malliaris & Salchenberger, 1996b).

We adopted the standard approach for training and testing, that is, to evaluate a model by testing its performance on a validation set consisting of out-of-sample data. Neural networks were trained on years 1997-1999 and forecasts generated on the validation set consisting of data for 2000.

Data Set Segmentation: Neural Clustering

The first set of neural networks was developed, as previously described, using the input set consisting of the price ratio data as described for the Australia, Hong Kong, and Japanese markets to predict the price ratio for the S&P 500 of the U.S. The results from using a training set consisting of daily observations for years 1997-1999 and a test set consisting of observations for 2000 were unsatisfactory. Further testing on networks developed, using prices divided into subsets based on quarters, did not yield significant improvement with respect to the performance measures selected for this problem.

Neural clustering was then employed to attempt to get better results with our neural network and to determine if there were some factors at work in this data set that were not obvious. The assumption was that better results (improved prediction accuracy) with the neural networks might be achieved when clustering was used to reduce the size of the input space and cluster the data into subsets that are highly correlated with each other and exhibit the same behavior.

In this study, two clustering techniques were used: k-means networks and Kohonen networks. K-means clustering is a fast-clustering technique that requires the number of clusters be selected in advance, and a minimum distance classifier is used to separate examples. Thus, an example is assigned to a cluster if it is closest to the center of that cluster. An initialization step, such as randomly assigning one example to each cluster is required to begin the process. Then each case is examined, distances are computed, and it is assigned to the cluster with the center closest to the case. After the case is assigned, the center of its cluster is updated. The process continues until all the examples are grouped into the specified number of clusters and further processing yields no change in cluster assignment.

Kohonen neural networks are unsupervised, self-organizing map (SOM) networks that project multidimensional points onto a two-dimensional network to simplify the complex patterns often found in high-dimensional input spaces. There are no middle layers, only input and output layers. These networks employ competitive learning and are useful in applications where it is important to analyze a large number of examples and identify groups with similar features. In competitive learning, "winner takes all" is used, where the "winner" is the connection with the highest firing rate. Units that are spatially close develop similar memory vectors, and neighborhoods shrink. Each cluster center can be thought of as describing the neural memory of a typical pattern for that cluster. Refer to chapter one of this text for additional discussion of self-organizing maps.

When an input pattern is presented, the units in the output layer compete with each other for the right to be declared the winner. The winner is the output node whose connection weight is the "closest" (mimimum distance) to the input pattern. The connection weights of the winner are adjusted and thus moved in the direction of the input pattern by a factor determined by the learning rate. As the process continues, the size of the neighborhoods decreases. Thus, the Kohonen network finds the closest matching neuron to the input and moves the weights of this neuron and those in neighboring proximity towards the input vector.

Clementine Data Mining software was used to conduct this cluster analysis. Clementine employs visual modeling techniques that allow the user to integrate preprocessing, model building, and evaluation of results by manipulating icons. In this research, we used the capabilities of this software package and Microsoft Excel® to eliminate outliers, consolidate the data, transform the data by applying mathematical functions to the data, establish clusters using two different techniques, develop, train, and test a neural network, and examine the results.

The actual determination of the number of clusters to be used is more of an art than a science. In Figures 1a to 1c, the results of clustering the data into 15, 9, and 6 clusters are shown. We want the number of clusters to be as small as possible while retaining the ability to distinguish between classes. Note that 9 clusters appear to give the best results, based on a visual inspection of the figures and the statistical output from Clementine. With 6 clusters, no distinct boundaries appear between clusters.

Figures 2a and 2b compare the results of using clustering on the training data set for 1997-1999 and the validation set. These figures show that the training and validation sets demonstrate similar clustering patterns. Where one cluster is sparsely populated in the training set, it is also sparsely

Figure 1a: Input data with 15 clusters

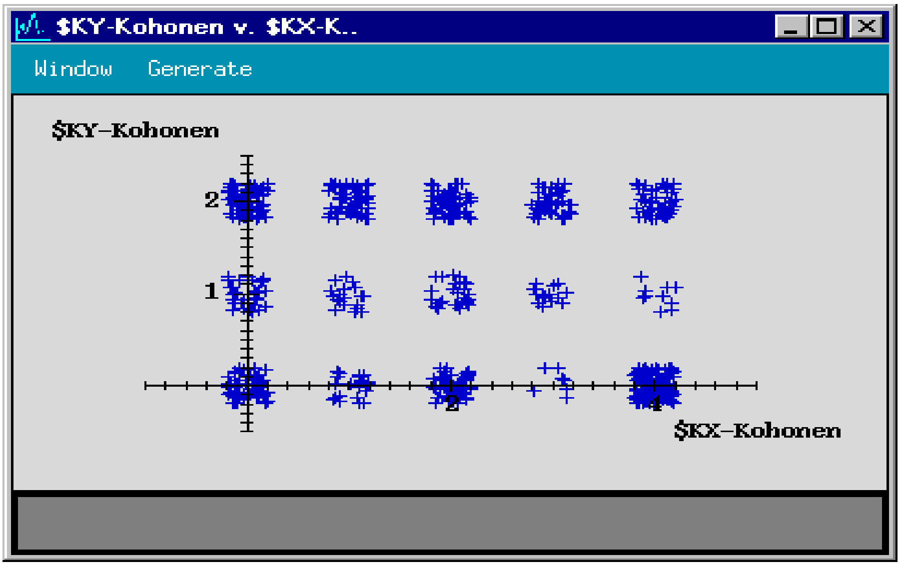
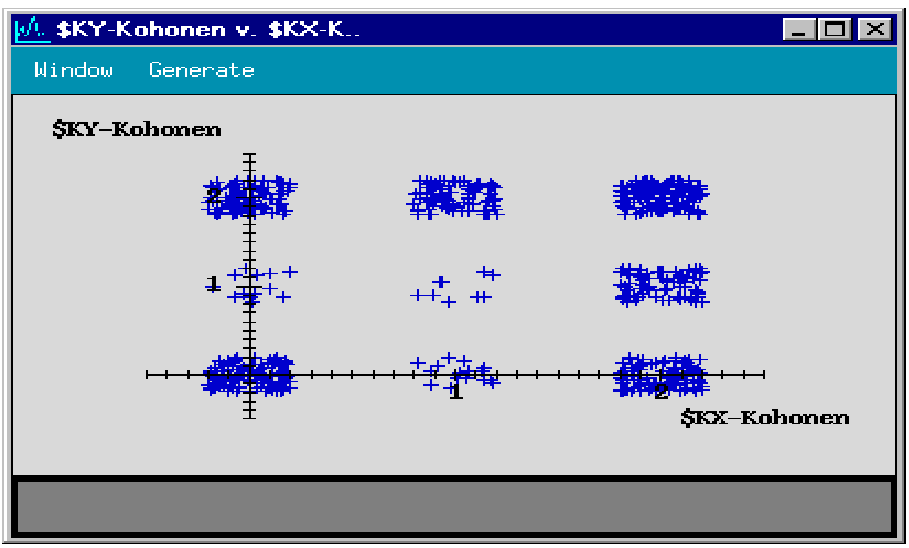


Figure 1b: Input data with 9 clusters



populated in the validation set. Coordinates as shown, from (0,0) to (2,2) were associated with each cluster and used to identify each cluster.

Having decided to use 9 clusters, we also developed clusters using the k-means technique. Nine clusters numbered 1-9 were formed for this analysis.

To determine if any improvement could be made to the neural network forecast, if the cluster association were known, two new training and test sets were developed. The first new data set added the k-means cluster value as an input to signal the cluster association to the network. A second data set included the Kohonen feature map cluster coordinate. Each of these sets was then used to train

Figure 1c: Input data with 6 clusters

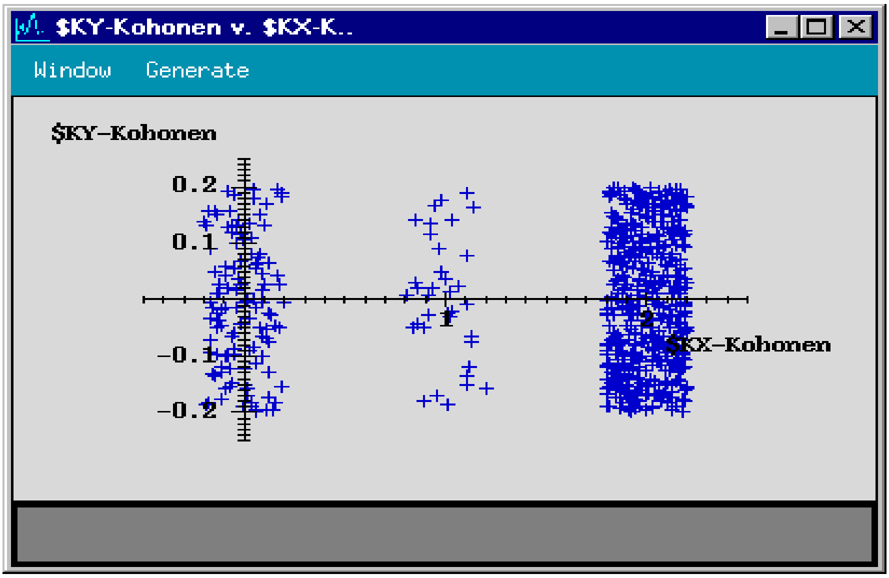
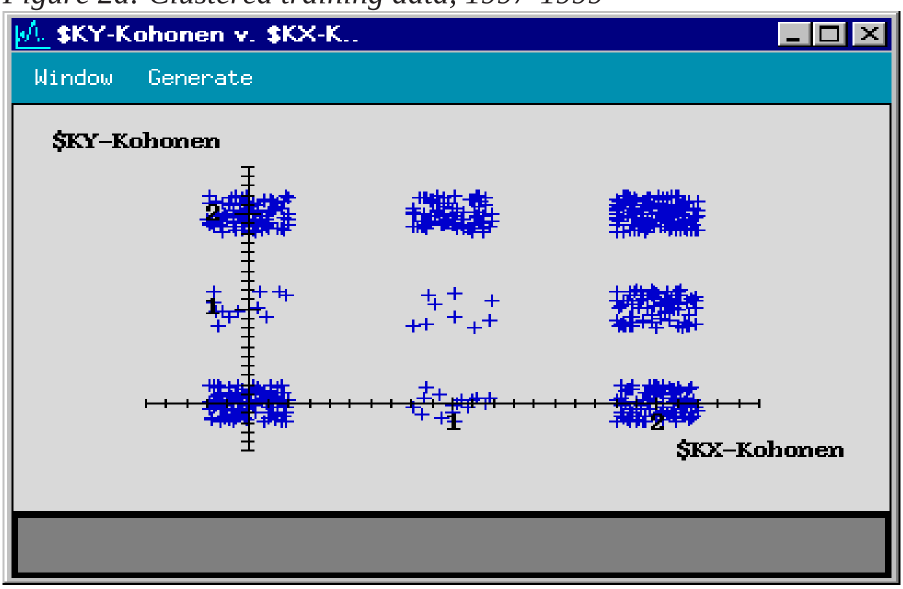


Figure 2a: Clustered training data, 1997-1999



a backpropagation neural network, and the validation set was used to generate preliminary results, as shown in Table 1. Only slight differences in the MSE for the prediction set were observed. We continued our experiments by using the clustering results in a different way.

Figure 2b: Clustered validation data, 2000

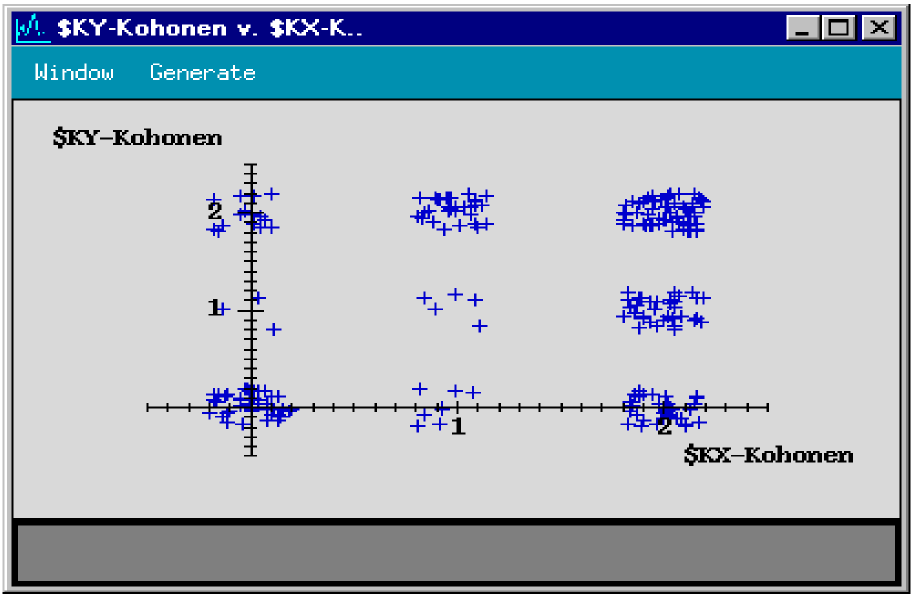


Table 1: Results without clustering

Model	MSE
Random Walk	0.0151
Neural Network	0.0106
Neural Network with Kohonen Variable Added	0.0107
Neural Network with K-Means Variable Added	0.0105

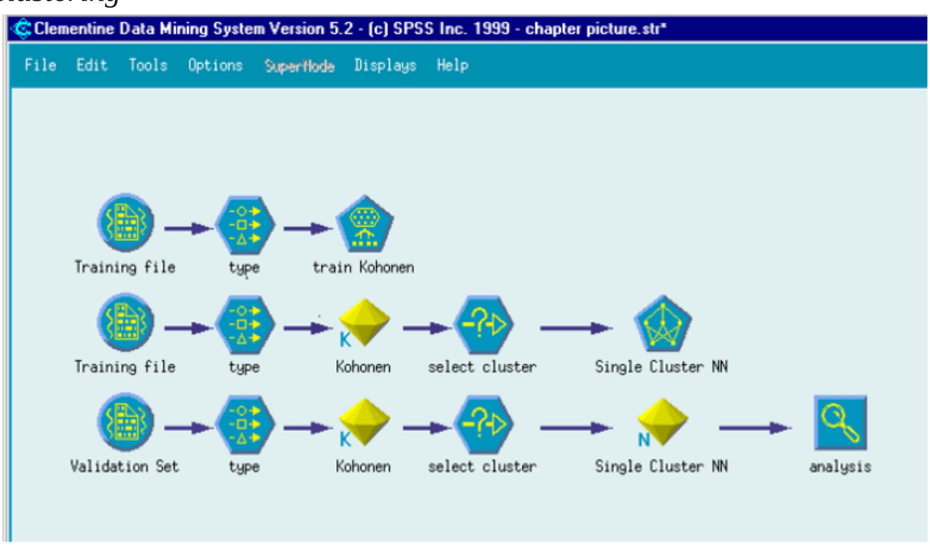
Developing the Neural Networks Using the Clustered Data Set

The real value of clustering the data is to discover data partitions that have not occurred to the human decision-maker. To see if this would improve the network's ability to develop daily forecasts, we next developed 18 separate neural networks, 9 for the k-means clusters and 9 for the Kohonen clusters. For each network, Clementine determines the best number of middle-layer nodes based on prediction accuracy of the training set, and the learning rate is initially set to 0.9.

Thus, instead of the single backpropagation neural network originally developed and tested, we needed to train and test a different neural network for each cluster, using the clusters discovered in the data set for the years 1997-1999. The first step was to train the 9 neural networks for the Kohonen clusters, and this was done using Clementine data mining software. Next, each observation in the validation set was fed into the trained Kohonen network and the resulting cluster value was identified. For example, if an observation from 2000 was most closely associated with the first cluster, then "0,0" was used to identify the cluster. Then the corresponding trained neural network was used for predicting the S&P price ratio. This process was repeated for the k-means clusters. For the validation set, the number of observations in each of the nine k-means clusters was 29, 2, 5, 41, 2, 26, 22, 14, and 98.

In Figure 3, the visual model developed using Clementine that trains and tests the appropriate networks is shown. The training set consisting of data from 1997 to 1999 is shown as the icon labeled "training file," and the validation set (data from 2000) is displayed in the model as the icon "validation set". The type icon is used by Clementine to identify the data type (e.g., integer, text, etc.) and purpose. The top row of the figure shows the process used to develop the Kohonen clusters. The center row shows the process of feeding the data through the trained Kohonen model. Then, a single cluster is isolated and a neural network is trained on that cluster. In the bottom row, data from the validation set are fed through the trained Kohonen model to identify the clusters. A cluster is selected and fed into the corresponding single-cluster, trained neural network. The results are then analyzed.

Figure 3: Neural network prediction model in Clementine, with Kohonen clustering



RESULTS

The initial results of developing backpropagation networks trained with data from 1997-1999 and tested using data from 2000, displayed in Table 1, led to the conclusion that better short-term predictions could be achieved if the data were properly clustered. The results of next set of experiments, using neural clustering, are shown in Tables 2and 3. In all cases, except one, the neural network based on a single cluster outperforms the random walk forecast for the same set of data. Further statistical analysis shows that these are significantly different at the.05 level of significance for the Kohonen clusters (0,0), (0,2), (1,0),(1,1),(1,2).

SUMMARY AND CONCLUSIONS

The results are interesting and significant from a methodological and an empirical perspective. The nature of the prediction problem and the results from the neural networks developed using the entire data set led to the decision to use two data mining tools for this prediction problem: clustering and neural networks. The quality of the data set in terms of its predictive capabilities often determines the success or failure of neural networks. This prompted the decision to reexamine the data set for strong relationships in the data set that we

Table 2: Random walk and neural network forecast errors within k-means clusters for year 2000

K-Means Cluster	MAE Random Walk	MAE Neural Network	RMSE Random Walk	RMSE Neural Network
1	0.0183	0.0107	0.0216	0.0136
2	0.0526	0.0203	0.0654	0.0247
3	0.0207	0.0114	0.0250	0.0159
4	0.0149	0.0115	0.0194	0.0151
5	0.0037	0.0139	0.0046	0.0139
6	0.0142	0.0099	0.0185	0.0115
7	0.0151	0.0125	0.0186	0.0166
8	0.0122	0.0079	0.0159	0.0090
9	0.0141	0.0098	0.0175	0.0134

Table 3: Random walk and neural network forecast errors within Kohonen clusters for year 2000

Kohonen Cluster	MAE Random Walk	MAE Neural Network	RMSE Random Walk	RMSE Neural Network
(0,0)	0.0163	0.0108	0.0196	0.0133
(0,1)	0.0172	0.0160	0.0230	0.0171
(0,2)	0.0149	0.0099	0.0221	0.0121
(1,0)	0.0153	0.0106	0.0190	0.0127
(1,1)	0.0171	0.0109	0.0213	0.0115
(1,2)	0.0205	0.0095	0.0218	0.0120
(2,0)	0.0117	0.0132	0.0141	0.0092
(2,1)	0.0158	0.0130	0.0192	0.0174
(2,2)	0.0162	0.0132	0.0201	0.0175

could not discover through knowledge of the problem domain that might be affecting the prediction results. The data was segmented into clusters based on features discovered through the clustering process, and the results were indeed improved.

The neural networks developed in this study outperformed the random walk predictions in most cases. This is an important empirical result, because the implication is that prices can be predicted using available information, thus signaling the existence of profitable trading strategies. That is, what happens daily in Japan, Australia, and Hong Kong does have an effect on the S&P 500, not only in catastrophic times, but also in normal day-to-day trading.

Using neural networks for price forecasting represents a valuable approach to this problem for several reasons. Neural networks may prove to be useful for these forecasting problems, which traditional statistical methods have been unable to solve. With neural networks, there is no need to engage in a debate over issues like autocorrelation, the probability distribution of the variables, or the nature of the underlying process, which must be determined before the statistical techniques traditionally used in futures prices forecasting can be used to develop forecasts. Since many of these issues appear to be unresolved, to the extent that conflicting evidence has been reported in many studies, a modeling approach which need not resolve these issues represents a great advantage.

The results of this study give us many leads for areas of future research. Both k-means and Kohonen clusters led to improved neural network forecasts, yet the clusters occurred in different ways. That is, the Kohonen and k-means clusters were not identical. Further analysis of the clusters – why they aggregate data into those groups and why the groups are not identical – is left to future analysis.

Since we have established that these markets have an effect daily, the door is now open to other researchers to investigate and refine these forecasts. The dominance of American markets is taken for granted. This study has shown that the major player is itself affected on a daily basis by the movements of markets on the opposite side of the globe. The world is indeed a small place.

REFERENCES

Azoff, E. M. (1994). *Neural Network Time Series Forecasting of Financial Markets*. Chichester, England: John Wiley & Sons Ltd.

Cavaletti, C. (1996). The who-what-where of linkages. *Futures*, 25(7), 62-65.

Dickinson, D. (2000). Stock market integration and macroeconomic fundamentals: An empirical analysis. *Applied Financial Economics*, 10(3), 261-276.

Lee S. B. and Kim, K. W. (1994). Does the October 1987 crash strengthen the co-movements among national stock markets? *Review of Financial Economics*, 3(2), 89-103.

Malliaris, A. and Urrutia, J. (1996). European stock market fluctuations: Short and long term links. *Journal of International Financial Markets, Institutions, and Money*, 6(2-3), 21-33.

Malliaris, M. and Salchenberger, L. (1993). A neural network for estimating options prices. *Journal of Applied Intelligence*, 3.

Malliaris, M. and Salchenberger, L. (1996a). Predicting wheat futures prices using neural networks. *International Journal of Computational Intelligence and Organizations*, 1(4).

Malliaris, M. and Salchenberger, L. (1996b). Using neural networks to forecast the S&P 100 implied volatility. *Neurocomputing*, 10.

Masih, A. and Masih, R. (1997). Dynamic linkages and the propogation mechanism driving major international stock markets: An analysis of the pre- and post-crash eras. *Quarterly Review of Economics and Finance*, 37(4), 859-886.

Nasseh, A. and Strauss, J. (2000). Stock prices and domestic and international macroeconomic activity: A cointegration approach. *Quarterly Review of Economics and Finance*, 40(2).

Theodossiou, P. (1994). Models for predicting prices and volatility patterns in major international stock markets. *Managerial Finance*, 20(5-6), 5-13.