Forecasting Futures Trading Volume Using Neural Networks

IEBELING KAASTRA
MILTON S. BOYD

INTRODUCTION

Neural networks are a type of artificial intelligence technology that attempt to mimic the human brain's powerful ability to recognize patterns. Typical applications in finance include mortgage risk assessment, risk rating of exchange-traded fixed income investments, portfolio selection/diversification, simulation of market behavior, index construction, and identification of economic explanatory variables [Trippi and DeSieno (1992)]. Applications also include forecasting financial time series such as stock and commodity prices [Trippi and Turban (1993)]. As universal function approximators, neural networks have the ability to handle nonlinear and chaotic time series data better than traditional forcasting models [Hornik et al. (1989)].

This article presents a neural network application in time series forecasting. Specifically, the objective of this article is to design neural networks to forecast monthly futures trading volume for the Winnipeg

Funding for this project was provided by the Winnipeg Commodity Exchange Fellowship.

- *Iebeling Kaastra is the Manager of Pricing at the Canadian Wheat Board.*
- *Milton S. Boyd is an Associate Professor in the Department of Agricultural Economics at the University of Manitoba.*

The Journal of Futures Markets, Vol. 15, No. 8, 953–970 (1995)

CCC 0270-7314/95/080953-18

Commodity Exchange (WCE). Commodities examined include barley, canola, flax, oats, rye, and wheat. Separate neural networks are designed for each monthly forecast up to nine months ahead. The forecasting power of the neural networks are evaluated using root mean squared error and mean absolute percentage error, and are then compared to the naive model using the Theil U statistic. The neural network's forecasting ability is also benchmarked relative to an ARIMA model.

Transaction fees on futures trading volume are an important component of the WCE's revenues. They accounted for 21% of the WCE's total annual operating budget in the fiscal year 1992/93. Canola accounted for 63% of all transaction fees over the same period. There are no formal models in place to forecast futures trading volume at the WCE. Current practice is to base trading volume forecasts on recent levels and keep any growth projections conservative.

Monthly mean absolute percentage changes over the past 16 years are large, ranging from 20% to 35% depending on the commodity, with standard deviations ranging from a low of 28% for canola and a high of 55% for rye. Such large fluctuations make forecasting trading volume (and the transaction fees derived from trading volume) difficult. Monthly pro-forma budgets require an accurate forecast of transaction fees to assist a commodity exchange in planning its operations.

This article is divided into four main sections. The first section briefly describes the structure and training of the popular backpropagation neural network which is used in this study. This is followed by a discussion of the forecasting evaluation methods and the data. The third section covers the design of the neural network forecasting models in more detail. It is arranged into four parts: variable selection, data preprocessing, neural network topology, and the training and forecasting procedure. The last section consists of an evaluation and comparison of the neural network models and the ARIMA models followed by a brief summary.

BACKPROPAGATION NEURAL NETWORKS

Backpropagation (BP) neural networks consist of a collection of inputs and processing units, generally known as either neurons, neurodes, or nodes (Fig. 1). The neurons in each layer are fully interconnected by connection strengths called weights which, along with the network architecture, store the knowledge of a trained network. In addition to the processing neurons, there is a bias neuron connected to each processing unit in the hidden and output layers. The bias neuron having a value of

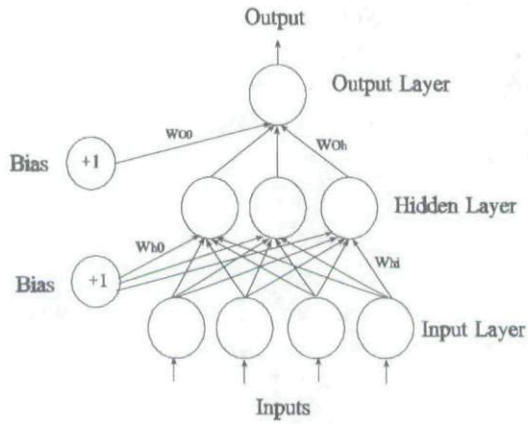


FIGURE 1
A feedforward neural network.

positive one is analogous to the intercept term in a regression equation and provides the network with more flexibility in classifying nonlinear problems. The neurons and bias terms are arranged into layers: an input layer, one or more hidden layers, and an output layer.

BP networks are a class of feedforward neural networks with supervised learning rules. Feedforward refers to the direction of information flow from the input to the output layer. Inputs are passed through the system once to determine the output. Supervised learning is the process of comparing each of the network's forecasts with the known correct answer and adjusting the weights based on the resulting forecast error. The number of input neurons is equal to the number of independent variables in the model while the output neuron(s) represent the dependent variable(s). The number of hidden layers and neurons within each layer can vary depending on the size and nature of the training set.

The BP training algorithm consists of a two-step process. First, inputs are passed forward to the first (or only) hidden layer. For each hidden layer neuron a weighted sum of the inputs is calculated by multiplying each input neuron's value by its respective weight. The weighted sum is modified by a transfer function which then passes the output forward to the next hidden layer or the output layer. A weighted sum is calculated and passed through its transfer function for each hidden layer and output layer neuron. In the second step, an error signal is calculated by comparing the forecasts at the output layer with the known correct answer. This error signal is "backpropagated" to the hidden layer(s) and the input layer with the weights being adjusted proportionally to each neuron's contribution to the forecast error (see the Appendix for a more detailed explanation of the BP algorithm).

The BP network is used in this study because it is the most common multi-layer network. Hornik et al. (1989) showed that the standard BP network using an arbitrary transfer function can approximate any measurable function in a satisfactory manner, if a sufficient number of hidden neurons is used. Hecht-Nielsen (1989) also demonstrated that a three-layer BP network can approximate any continuous mapping. Details of neural network procedures and backpropogation are shown in the Appendix and a complete discussion of neural network learning can be found in White (1992).

FORECASTING EVALUATION METHODS

Three criteria are used to evaluate and compare the forecasting power of the neural networks and the ARIMA models by commodity and by forecast horizon. The first is the mean absolute percentage error (MAPE). It is a measure of the average absolute error for each point forecast made by the neural networks. Dividing each absolute error by its actual observation allows comparisons of forecasting accuracy to be made among the six commodities. MAPE is given by

$$\mathrm{MAPE} = \frac{\Sigma |(P_t - A_t) \div A_t|}{N}$$

$$t = t + 1, t + 2, \ldots, N \tag{1}$$

where P and A are the predicted and actual values, respectively.

The second criterion is the root mean squared error (RMSE) which measures the overall performance of a model. It is a commonly used measure of forecast error and goodness-of-fit. The formula for RMSE is

$$\mathrm{RMSE} = \sqrt{\frac{\Sigma (P_t - A_t)^2}{N}} \tag{2}$$

The third criterion is the naive forecast. The simplest naive forecast is $P_t = A_{t-1}$. This forecasting standard is implicit in the coefficients of inequality (U_2) proposed by Theil (1966), defined as follows:

$$U_2 = \frac{\sqrt{\frac{\Sigma (P_i - A_i)^2}{N}}}{\sqrt{\frac{\Sigma A_i^2}{N}}} \tag{3}$$

where $P_i = P_t - A_{t-u}$ and $A_i = A_t - A_{t-u}$, where u is the number of step-ahead forecasts. The larger the value of U_2, the worse the model's forecasting ability. Perfect forecasts are implied when U_2 is equal to zero. If U_2 is equal to one that is equivalent to the forecasting ability of a naive model which used only last period's actual value of the dependent variable. The forecasting accuracy of the neural network and the ARIMA models are worse than the naive model if U_2 is greater than one.

DATA

Total monthly transactions for barley, canola, flax, oats, rye, and wheat measured in 20 tonne contracts are used to train the neural networks and estimate the ARIMA models.[1] Each commodity's total trading volume is the sum of the transactions of all the futures delivery months traded during that month. Data for the crop years 1977/78 to 1987/88 are from various issues of the Winnipeg Commodity Exchange (WCE) Statistical Annual.[2] The internal records of the WCE provided total trading volume data from 1988/89 to 1992/93 due to the change in reporting in 1988/89. Open interest is the daily average open interest of all futures delivery months as reported in the WCE Statistical Annuals. Monthly mean cash price in dollars per metric tonne is the daily average close of cash transactions at the WCE obtained from various issues of the Grains Industry Handbook published by the Canada Grains Council. The grade of the cash commodity chosen is based on the grade deliverable at par on the respective futures contract. The futures price series consists of the nearby futures closing prices obtained from the data vendor Technical Tools. Total producer deliveries to licensed elevators in the provinces of Alberta, Saskatchewan, and Manitoba are from the Canadian Grain Commission's Visible Supplies and Disposition publication. The entire time series consists of 192 observations.

VARIABLE SELECTION

Five monthly independent variables are chosen as likely predictors of futures trading volume. These include lagged trading volume, open inter-

[1] Prior to the 1988/89 crop year, volume of trade published in the Winnipeg Commodity Exchange Statistical Annual included the following transactions: all broker allocations to principals, give-ups, error corrections, and exchanges of futures for physicals. These transactions are no longer included in the volume figures. Since transaction fees received by the WCE are based on total transactions, which include both long and short trades, it is more appropriate for the neural networks to be forecasting total transactions for budgeting purposes rather than the recently revised trading volume figures.

[2] A crop year is from August 1 to July 31.

est, futures price variability, mean cash prices and producer deliveries to licensed elevators in the three prairie provinces. Each of these variables is preprocessed before being fed into the neural network as an input. (This is explained later.)

Open interest reflects the average daily outstanding long or short positions during the month and can be interpreted as a measure of the expected future positions of hedgers since intra-day speculators do not affect open interest. Martell and Wolf (1987) found a statistically positive correlation between monthly open interest and futures trading volume for five metals traded on the COMEX. This is not surprising considering that all other things being equal, volume will always increase whenever open interest changes. Futures trading volume and one-month lagged open interest for all six commodities on the WCE are positively correlated and together could be useful predictors in a neural network forecasting model.

Price volatility has been shown to be positively related to futures trading volume. Karpoff (1987) in a survey paper found that 18 out of 19 studies investigating the relationship between absolute price change and volume report a positive correlation. Price volatility and volume traded should be positively related because of the joint dependence on a common underlying variable which can be interpreted as the rate of information flow to the market [Najand and Yung (1991)]. Price volatility is measured using the standard deviation of daily nearby futures prices. Hedgers would likely increase trading in response to increasing price volatility to protect against potentially large cash market losses. Speculators may be attracted also to the futures market because the opportunities for profit are increased with increasing price volatility.

Another explanation of the relationship between price volatility and trading volume is the expectational heterogeneity-volume model [Trippi and Harriff (1993)]. This assumes that the relationship between heterogeneity of risk expectations and volume is caused by changing levels of different assessments of risk between asset holders and nonasset holders. To equate the various changing preferences between agents, trading must occur. In this case, it may be possible that risk can be described by changing nonconstant variance models which are nonlinear, such as autoregressive conditional heteroskedasticity (ARCH) models. If this is the case, then more generalized nonlinear models such as neural networks may be suitable for modeling rather than traditional linear models.

Mean cash prices are included as inputs since short hedgers may believe that, when mean cash prices are at historic lows, there is little need to hedge. Conversely, long hedgers may be reluctant to hedge when prices are near their highs.

Producer deliveries are likely a good proxy for hedging demand by grain handling companies and exporters. Hedging demand is very important at the WCE since it is largely a commercial market where hedging volume accounts for the majority of total trading volume.

DATA PREPROCESSING

The design of a reliable neural network forecasting model requires considerable experimentation. The selection of proper inputs, as in any forecasting model, is likely of greatest importance. The transformation of raw data to minimize noise, highlight important relationships, detect trends, and flatten the distribution of the variable to assist the neural network in learning the relevant patterns is called preprocessing.

Preprocessing of the training and testing sets indicates that higher correlations can be obtained by presenting the neural network with a three-period arithmetic moving average of the inputs. The moving averages smooth the data to minimize noise, detect trends, and may assist in capturing some seasonality. The use of logarithmic transformations on all variables and the use of a seven-period moving average prove to be less successful. The three-period moving average is applied to all inputs for each of the six commodities.

All variables are scaled between zero and one using mean/standard deviation scaling. Observations plus or minus two standard deviations from the mean are mapped to one and zero, respectively. This type of scaling results in a more uniform distribution and helps to suppress the effect of outliers. Given the small training set ($n = 148$) and the scaling method used, no outliers are removed from the training and testing sets.

All inputs, except for producer deliveries, are lagged from one to ten months depending on the number of periods the network must forecast ahead. Producer deliveries are lagged an additional month to account for the fact that the information is not publicly released until the middle or end of the following month.

NEURAL NETWORK TOPOLOGY

Chakraborty et al. (1992) classify neural networks as a data-driven approach rather than a model-driven approach. Neural networks look for patterns in the training set, learn these patterns, and develop the ability to correctly classify new patterns and make forecasts. The ability of a neural network to learn the patterns in the data is determined in large part by the number of weights interconnecting the neurons in each of the layers. A network with too few weights is unable to learn the patterns, while a network with too many weights will overfit the data [Masters (1993)]. Overfitting occurs when a model has too few degrees of freedom relative to its parameters.

In the case of neural networks, the number of weights relative to the size of the training set determine the likelihood of overfitting. A neural network with too many hidden neurons will memorize and overfit each training example and the network will fail to predict on out-of-sample data [Bowen (1991)]. In this study, the number of input neurons is fixed at five for all neural networks. That number should balance the need for a sufficient set of independent variables, while limiting the number of weights in the neural network to avoid overfitting.

The number of hidden neurons in the single hidden layer are selected experimentally based on the testing set performance of each neural network. Three, five, seven, and nine hidden neurons are evaluated, resulting in a total of 22, 50, and 64 weights, respectively, including bias terms and output neurons. Networks with more than nine hidden neurons are not evaluated because overfitting is very likely to occur when the number of training facts relative to the number of weights approaches two to one. Increments of two for the hidden neurons allow a reasonable range to be tested while keeping computational times feasible.

One neural network is designed for each of the nine monthly forecasts. The output neuron used in each neural network represents the trading volume for one to nine months ahead. For example, one neural network forecast next month's trading volume while another neural network specializes in forecasting the following month. In all, nine neural networks are trained for each commodity with each differing only by the number of step-ahead forecasts.

A three-layer, fully interconnected backpropagation neural network is used in this study. Transfer functions for the input neurons are linear, while all others are sigmoid. Each neural network is trained to minimize the total squared error of all training facts. The choice of transfer and error functions is consistent with a standard BP neural network.

TRAINING AND FORECASTING PROCEDURE

The multidimensional error surface present in multi-layer neural networks often contains many local minima in which the training algorithm can become stuck. Local minima are believed to be especially troublesome for networks with a small number of hidden neurons, as is the case here [Masters (1993)]. Training each neural network multiple times by using randomly selected starting weights is one common method to increase the likelihood of not becoming trapped in a local minimum. In this study, each neural network is trained seven times for 5000 iterations using randomly selected starting weights which should provide for a reasonable opportunity of reaching a global minimum.

The network with the highest mean correlation on the testing set, out of the seven random sets of starting weights, is saved by the neural network. The testing set is randomly selected from the first 168 observations (14 crop years). Once selected, the 20 testing observations are removed from the training set, leaving 148 training facts. Randomly selecting testing observations is used to insure that the testing set contains both volume decreases and increases of varying magnitudes. In this way, better out-of-sample performance is possible.

The neural network with the lowest MAPE on the testing set for each commodity is used to forecast futures trading volume on the validation set. The validation set consists of the 24 observations from the crop years, 1991/92 and 92/93. In situations where the forecasting performance of the testing set is similar between two networks, the network with the smallest number of hidden neurons is selected.

The nine separate forecasts for each of the six commodities, tested over four different hidden neurons, with seven sets of random starting weights, results in a total of 1512 networks that require training. Training time for all six commodities is approximately 110 hours on a 486-DX2 66 Mhz. computer using the neural network software NTRAIN from Scientific Consultant Services. Extensive use of its batch/script file capability is made to automate the training and testing procedure.

EVALUATION AND COMPARISON

Overall, the results for the neural networks are mixed with forecasting performance on some commodities much better than others. Table I and II present the Theil U, RMSE, and MAPE statistics for the six commodities examined over a nine-month forecast horizon for the neural network and the ARIMA models, respectively. Since an ARIMA model can be estimated as an AR model if the MA process is invertible, and since AR models are easier to estimate, have well-developed model selection criteria, and require limited pretesting, they are the form of ARIMA estimated here. The typical AR estimation procedures are used

TABLE I

Results of Out-of-Sample Futures Trading Volume Forecasts by Neural Network Models, Winnipeg Commodity Exchange, 20 Tonne Contracts 1991/92–1992/93

Forecast	RMSE[a]						MAPE[b]						Theil U					
Horizon	Barley	Canola	Flax	Oats	Rye	Wheat	Barley	Canola	Flax	Oats	Rye	Wheat	Barley	Canola	Flax	Oats	Rye	Wheat
1 month	5,891	23,135	7,947	4,098	2,908	8,107	37.4	13.1	34.3	53.3	52.4	19.0	1.09	0.82	0.89	1.17	1.13	0.85
2 months	5,879	31,246	8,324	4,782	2,918	8,214	36.9	18.9	33.7	58.0	58.5	19.3	1.18	0.98	0.96	0.98	1.14	0.86
3 months	9,188	24,611	7,190	3,864	3,896	9,031	65.6	15.4	30.9	52.5	155.4.	18.6	1.82	0.81	0.88	0.73	1.55	0.94
4 months	6,796	32,487	6,455	3,317	2,949	6,603	42.8	22.5	29.5	49.1	98.5	18.0	1.30	0.88	0.69	0.65	1.02	0.66
5 months	6,728	27,471	7,712	4,101	3,137	7,043	43.5	17.2	35.0	44.3	124.9	19.0	1.46	0.89	1.08	0.94	1.37	0.67
6 months	9,213	31,241	11,328	3,119	3,648	7,553	58.5	20.9	51.2	47.2	70.3	18.4	1.45	0.83	1.23	0.69	1.36	0.73
7 months	13,065	25,076	9,807	4,146	3,124	9,403	89.2	15.7	44.9	58.9	91.9	25.0	2.85	0.64	1.17	0.89	1.20	1.15
8 months	16,374	30,361	7,110	5,213	2,579	8,255	113.3	19.3	33.5	70.3	56.1	20.3	2.87	0.90	0.85	0.95	0.83	0.89
9 months	15,111	27,933	9,248	5,010	3,016	9,303	107.5	19.4	42.8	67.1	119.2	24.7	2.39	0.78	1.10	0.83	1.13	0.76
Mean	9,805	28,173	8,347	4,183	3,131	8,168	66.1	18.0	37.3	55.6	91.9	20.3	1.82	0.84	0.98	0.87	1.19	0.83

[a]Root mean squared error.
[b]Mean absolute percentage error.

TABLE II

Results of Out-of-Sample Futures Trading Volume Forecasts by the ARIMA Models, Winnipeg Commodity Exchange, 20 Tonne Contracts 1991/92–1992/93

Forecast	RMSE[a]						MAPE[b]						Theil U					
Horizon	*Barley*	*Canola*	*Flax*	*Oats*	*Rye*	*Wheat*	*Barley*	*Canola*	*Flax*	*Oats*	*Rye*	*Wheat*	*Barley*	*Canola*	*Flax*	*Oats*	*Rye*	*Wheat*
1 month	5,421	25,639	7,013	4,409	3,276	7,064	34.1	18.6	31.2	68.3	115.7	22.5	0.98	0.87	0.78	1.10	1.23	0.94
2 months	5,873	41,134	7,344	4,946	4,183	9,256	38.3	24.9	29.6	70.0	151.9	26.3	1.19	1.30	0.87	0.94	1.75	0.97
3 months	5,582	45,963	8,277	5,759	4,583	9,564	38.6	27.0	35.4	79.5	169.3	25.6	1.09	1.51	1.04	1.08	1.71	0.94
4 months	5,973	52,112	8,648	6,600	4,793	10,103	39.8	35.4	37.0	87.8	179.6	27.9	1.27	1.42	0.94	1.19	1.70	0.95
5 months	5,737	71,309	10,877	7,319	4,823	10,652	39.8	43.3	49.9	98.1	181.4	29.0	1.47	2.31	1.51	1.41	2.06	0.96
6 months	5,736	99,808	11,917	7,841	4,870	10,750	38.6	62.3	47.2	104.8	183.4	29.4	0.94	2.66	1.30	1.60	1.85	1.02
7 months	5,848	114,221	14,194	8,140	4,879	11,318	37.6	77.7	60.1	106.5	184.0	32.1	1.28	2.94	1.69	1.72	1.90	1.30
8 months	5,958	169,316	13,835	8,267	4,881	11,155	37.7	122.4	49.1	106.9	183.8	30.5	1.07	5.06	1.69	1.44	1.64	1.17
9 months	15,111	234,718	18,630	8,426	4,899	10,516	38.5	166.1	79.8	109.8	184.8	28.8	1.04	6.40	1.91	1.38	1.57	1.22
Mean	6,804	94,913	11,193	6,856	4,576	10,042	38.1	64.1	46.6	92.4	152.0	28.0	1.15	2.72	1.12	1.32	1.71	1.05

[a]Root mean squared error.
[b]Mean absolute percentage error.

with the AIC method (Akaike) used to identify the lag length, and the Ljung–Box statistic used to confirm white noise in the residuals.

The Theil U statistic compares the neural network forecasting power relative to a naive model which uses only last period's trading volume as its independent variable. The neural networks generally outperform the naive model. Thirty-six of the 54 (67%) Theil U values in Table I are less than one. The ARIMA model outperforms the naive model in 12 out of 54 cases (22%). However, barley and rye are notable exceptions. The Theil U statistic is greater than one in 17 out of 18 cases. One possible explanation for the poor performance of both barely and rye is the large variability during the training and testing period as compared to very little fluctuations for all 24 out-of-sample observations. For example, barley trading volume during the training period regularly ranged from 20,000 to over 100,000 contracts per month (Fig. 2). During the validation period, trading volume ranged from 9,000 to 25,000 contracts and was in a general decline. Outliers were not removed from the training and testing sets and this may in part account for forecasts that are consistently too high. The same argument can be made for rye.

The forecasting performance of the neural networks relative to the naive model remains strong throughout the nine-month forecast horizon and does not appear to deteriorate as the forecast horizon increases. One possible explanation for this phenomenon is that the naive model forecasts are deteriorating while the neural network's nearby forecasts are not as good. The naive model is expected to perform well for the one-step ahead forecast as last period's value is likely a very good indicator of next period's trading volume. The neural networks, by using a three-period moving average, may lose out on the one-step ahead forecast, but gain accuracy for the longer term forecasts. The ARIMA model generally performs best for the nearby forecasts, but looses accuracy as the forecast horizon increases.

The MAPE vary widely by commodity and by forecast horizon. Canola and wheat exhibit the lowest MAPE, ranging from 13% to 25%, while barley and rye show the highest values. The strong forecasting performance for canola and wheat is likely attributable to the relatively low mean absolute percentage changes from month to month and the consistency in volume variability from the training/testing sets to the validation set. The RMSE follows a similar pattern as MAPE although there are instances in which the two statistics do not move in the same direction.

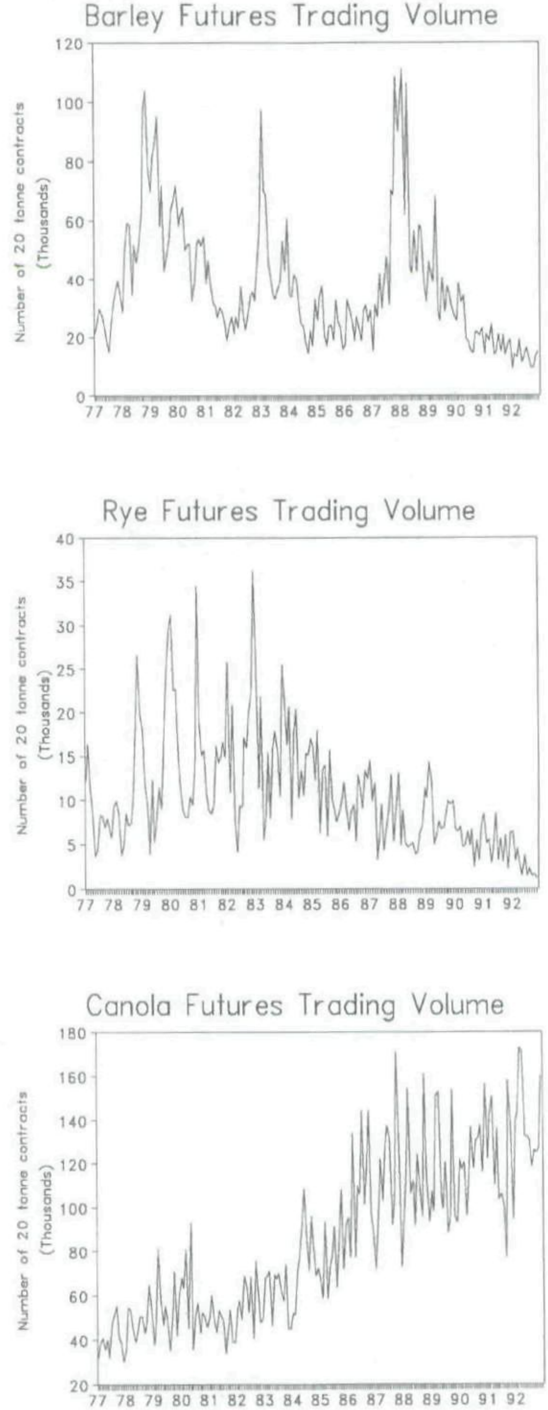
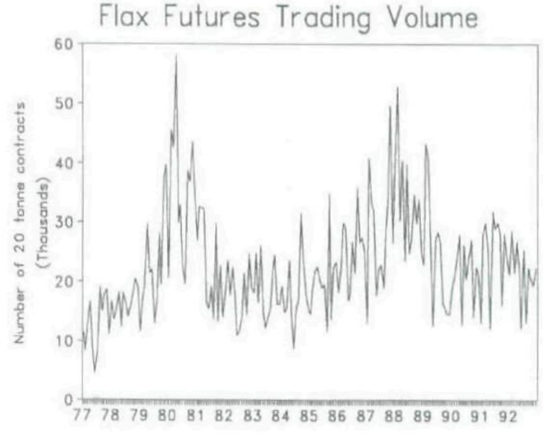
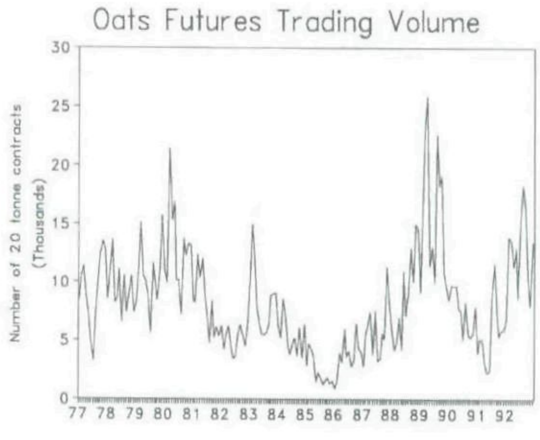


FIGURE 2

Futures trading volume at the Winnipeg Commodity Exchange, 1977/78–1992/93.





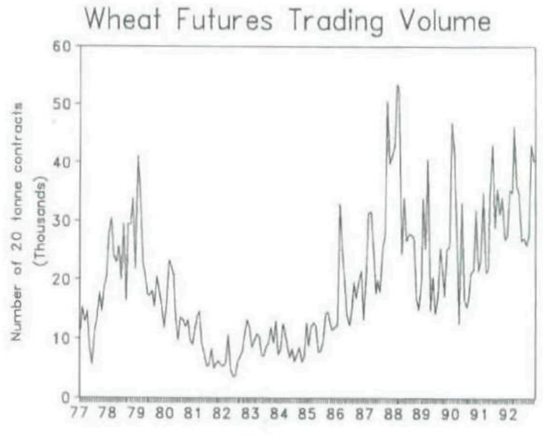


FIGURE 2 (Continued)

In terms of the network topology, the number of hidden neurons also vary by commodity and by forecast horizon. Table III shows that, out of the 54 networks listed, 23, 15, and 16 use three, five, and seven hidden neurons, respectively. Nine hidden neurons are not able to perform well on the testing set which is likely indicative of some overfitting during training. The strong showing of networks with three hidden neurons indicates that a small neural network can approximate the function quite well. However, a range of hidden neurons should still be tested because the network topology varies considerably by commodity and by forecast horizon.

Results show a relatively strong forecasting performance by the neural networks for canola and wheat. These two commodities combined accounted for over 76% of total transaction fees during 1991/92 and 92/93. On the other hand, the poor forecasting performance for rye and barley is less of a problem because together these two commodities accounted for slightly over 9% of total transaction fees during the same period. Also, the results are encouraging because the forecasts do not appear to deteriorate as the forecast horizon increases as does the naive model which is currently in place at the WCE.

SUMMARY

This study attempts to forecast futures trading volume on out-of-sample data. The backpropagation neural network is used to forecast futures trading volume for six commodities traded on the Winnipeg Commodity Exchange. The forecasts are designed to improve the budgeting of transaction fees. Neural networks are used because of their ability

TABLE III

Number of Hidden Neurons Selected in Final Neural Network Forecasting Models Based on Testing Set Performance

Forecast Horizon	*Barley*	*Canola*	*Flax*	*Oats*	*Rye*	*Wheat*
1 month	7	7	3	3	7	7
2 months	3	7	3	3	7	7
3 months	5	3	3	5	5	5
4 months	7	3	3	3	5	7
5 months	7	3	5	3	7	7
6 months	3	3	3	3	7	7
7 months	3	5	5	3	5	5
8 months	3	5	3	5	5	7
9 months	3	3	5	3	5	7

to approximate any continuous function and learn nonlinear patterns that may be present in the data and incorporate them into forecasts.

Separate neural network forecasting models are created for barley, canola, flax oats, rye, and wheat. Trading volume is forecasted up to nine months ahead with a single neural network specializing in each step-ahead forecast. The five independent variables used as inputs to the neural networks include: three-period moving averages of lagged trading volume, open interest, futures price variability, mean cash prices, and producer grain deliveries. Three, five, seven, and nine hidden neurons are trained and evaluated on the testing sets.

The results indicate that the neural networks are able to forecast up to nine months ahead and outperform the naive model for all commodities except barley and rye. The best forecasts are obtained for canola and wheat which together account for over three-quarters of total trading volume and are therefore the most important commodities for budgeting purposes. The neural network forecasts relative to the naive model currently in place to forecast futures trading volume at the WCE do not deteriorate as the forecast horizon increases. The neural network also outperforms an ARIMA model.

In terms of network topology, neural networks with three hidden neurons are selected for 23 out of the 54 networks. Networks with five and seven hidden neurons are selected almost equally often, while no network with nine hidden neurons performs well on the testing set.

APPENDIX

Standard BP employs an optimization method called gradient descent algorithm which follows the contours of the error surface by always moving down the steepest slope. The objective of training is to minimize the total squared errors, defined as follows [McClelland and Rumelhart (1986)]:

$$E = \frac{1}{2}\sum_{h}^{M} E_h = \frac{1}{2}\sum_{h}^{M}\sum_{i}^{N}(t_{hi} - O_{hi})^2 \qquad (1)$$

where E is the total error of all patterns; E_h represents the error on pattern h; the index h ranges over the set of input patterns; and i refers to the ith output neuron. The variable t_{hi} is the desired output for the ith output neuron when the hth pattern is presented, and O_{hi} is the actual output of the ith output neuron when pattern h is presented. The learning rule to adjust the weight between neuron i and j is defined as:

$$\delta_{hi} = (t_{hi} - O_{hi})O_{hi}(1 - O_{hi}) \tag{2}$$

$$\delta_{hi} = O_{hi}(1 - O_{hi})\sum_{k}^{N} \delta_{hk}w_{jk} \tag{3}$$

$$\Delta w_{ij}(n + 1) = \varepsilon(\delta_{hi} - O_{hj}) \tag{4}$$

where n is the presentation number; δ_{hi} is the error signal of neuron i for pattern h; and ε is the learning rate. The learning rate is a constant of proportionality which determines the size of the weight changes. The weight change of a neuron is proportional to the impact of the weight from that neuron on the error. The error signal for an output neuron and a hidden neuron are calculated by eqs. (2) and (3), respectively.

One method to increase the learning rate and thereby speed up training time without leading to oscillation is to include a momentum term in the backpropagation learning rule. The momentum term determines how past weight changes affect current weight changes. The modified BP training rule is defined as follows:

$$\Delta w_{ij}(n + 1) = \varepsilon(\delta_{hi}O_{hj}) + \alpha\Delta w_{ij}(n) \tag{5}$$

where α is the momentum term, and all other terms as previously defined.

The momentum term suppresses side-to-side oscillations by filtering out high-frequency variations. Each new search direction is a weighted sum of the current and the previous gradients. Such a two-period moving average of gradients filters out rapid fluctuations in the learning rate. Momentum values that are too great will prevent the algorithm from following the twists and turns in weight space while learning rates that are too small will require an inordinate amount of time to converge to a solution.

BIBLIOGRAPHY

Akaike, H. (1976): "Canonical Correlations Analysis of Time Series and the Use of an Information Criterion," in *Advances and Case Studies in System Identification,* Mehra, R., and Laniotis, D. G. (eds.), New York: Academic Press.

Bowen J. E. (1991): "Using Neural Network Nets to Predict Several Sequential and Subsequent Future Values from Time Series Data," *The First International Conference on Artificial Intelligence Applications on Wall Street,* IEEE, 4:30–34.

Caudill, M. (1992): "The View from Now," *AI Expert,* June:24–31.

Chakraborty, K., Mehrotra, K., Mohan, C. K., and Ranka, S. (1992): "Forecasting the Behaviour of Multivariate Time Series Using Neural Networks," *Neural Networks,* 5:961–970.

Hecht-Nielson, R. (1991): *Neurocomputing.* Menlo, CA: Addison-Wesley.

Hornik, K., Stinchcombe, M., and White, H. (1989): "Multilayer Feedforward Networks are Universal Approximators," *Neural Networks,* 2:359–366.

Karpoff, J. M. (1987): "The Relation Between Price Changes and Trading Volume: A Survey," *Journal of Financial and Quantitative Analysis,* 22:109–126.

Martell, T. F., and Wolf, A. S. (1987): "Determinants of Trading Volume in Futures Markets," *The Journal of Futures Markets,* 7:233–244.

Masters, T. (1993): *Practical Neural Network Recipes in* C++. New York: Academic Press.

McClelland, J. L., Rumelhart, D. E., and the PDP Group. (1986): *Parallel Distributed Processing. Explorations in the Microstructure of Cognition, Vol. I: Foundations.* Cambridge, MA: MIT Press.

Najand, M., and Yung, K. (1991): "A GARCH Examination of the Relationship Between Volume and Price Variability in Futures Markets," *The Journal of Futures Markets,* 11:613–621.

Scientific Consultant Services, Inc. Ntrain V1.02. 20 Stagecoach Road, Seleden, New York, 11784.

Theil, H. (1966): *Applied Economic Forecasting.* New York: Rand McNally.

Trippi, R., DeSieno, D. (1992): "Trading Equity Index Futures With a Neural Network," *Journal of Portfolio Management,* 19:27–33.

Trippi, R., and Harriff, R. (1993): "Trading Volume, Heterogeneity of Expectations, and the Dispersion of Volatilities by Option Prices," *Review of Quantitative Finance and Accounting,* 3:339–351.

Trippi, R., and Turban, E. (1993): *Neural Networks in Finance and Investing.* Chicago: Probus Publishing.

White, H. (1992): *Artificial Neural Networks: Approximation and Learning Theory.* New York: Blackwell Publishers.