FORECASTING AND SELLING FUTURES USING ARIMA MODELS AND A NEURAL NETWORK

by

Gordon Holens

A Thesis
submitted to the Faculty of Graduate Studies
in partial fulfilment of the requirements
for the degree of

MASTER OF SCIENCE

Department of Agricultural Economics and Farm Management
University of Manitoba
Winnipeg, Manitoba

January, 1997

0-612-23343-X

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION PAGE

FORECASTING AND SELLING FUTURES USING ARIMA MODELS AND A NEURAL NETWORK

BY

GORDON HOLENS

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirements of the degree

of

MASTER OF SCIENCE

ABSTRACT

This study involves comparing the forecasting and trading performance of an ARIMA model and a neural network model. The optimal ARIMA model is selected by choosing the combination of sample size and forecast ahead period that produce the minimum forecast error. Weekly data for two contracts traded on the futures exchanges are used. Results suggest that a mid range sample size together with the minimum forecast ahead period produces the lowest forecast error. Secondly, a neural network using the optimal sample size and forecast ahead period chosen from above is compared to the ARIMA model. It turns out that the neural network is able to lower the forecast error. This study also checks for the ability of both the ARIMA and neural network models to detect turning points in the market. It turns out that both models for both commodities are able to predict turning points with about the same degree of accuracy.

Lastly, the optimal ARIMA model together with the neural network model are used to trade futures contracts using a given trading strategy. The models all produce negative profits but the neural network suffers smaller losses per trade and trades slightly more often. Neither the neural network or the ARIMA models were able to sell at a significantly higher price than the overall average selling price. Overall, the negative profits produced by the models together with the low percentage of profitable trades may indicate that the trading regime is not appropriate. It may also suggest that the neural network is over fitting the data or that the ARIMA model is not well specified

ACKNOWLEDGEMENTS

I would like to extend my greatest appreciation to my advisor Dr. Milton Boyd for introducing me to neural networks and for his continuing invaluable support and enthusiasm. I would also like to sincerely thank Dr. Bruce Johnston for his helpful statistical advice. In addition, I would like to thank Dr. Darryl Kraft who provided guidance while Dr. Boyd was on sabbatical. Chrispin Ntungo also deserves recognition as he was extremely helpful with the data analysis. As well, I would like to express my gratitude to the Canadian Wheat Board and Manitoba Agriculture for their information and funding support.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Problem Statement

Over the last several years there has been an ongoing effort to model and forecast futures prices. Traditionally linear ARIMA time series models have been utilized to perform this task. They may be inappropriate due to such things as the non-linear behavior of price variables. For example Kohzadi and Boyd (1995) showed that there was nonlinear dynamics in cattle prices. In addition two other studies Blank (1991) and Chavas and Holt (1991) demonstrated nonlinearity in futures prices. Recently though there has been an increasing interest in the application of neural networks for forecasting futures prices. They have demonstrated in a number of studies to be able to outperform ARIMA models. For example, Kohzadi (1994) found that neural networks could outperform ARIMA models in terms of forecasting accuracy for U.S. cattle futures prices. In another study, Kohzadi and Boyd (1995) found neural networks produced lower forecast error than ARIMA models for corn futures prices when using a sliding window procedure.

Research Objectives

The objective of this study is to investigate three different aspects of a commodity price series. Firstly the study will look exclusively at ARIMA time series models. A set of models derived from a preset combination of various sample sizes and forecast periods will be estimated and used for forecasting. Each combination will yield a forecast error value which will be used for further analysis. The study will attempt to demonstrate if either changing the sample size or the forecast period affects the size of the forecast error. A factorial design will be used to analyze these results. Secondly the "best" ARIMA model from each of the two commodities will be used to compare its forecast error measure against the performance of a neural network. Finally both the ARIMA model and the neural network will be used to trade futures over a given time period to see whether positive profits can be generated for the two chosen commodities.

Hypotheses

It is expected from the research, that in general, larger sample sizes and shorter forecast periods will produce the lowest values for the forecast error. This is a new area of study which should be helpful in choosing the most appropriate ARIMA model. Clearly a researcher is interested in the best combination of sample size and forecast period to produce a minimum error. In addition, it is expected that the neural network will have lower forecast error than the chosen ARIMA model due to the non-linear behavior of most commodity price series. The third hypothesis is that the neural network should produce higher profits for both of the contracts.

Outline of Thesis

Following the introduction is two chapters, a reference section and an appendix. Chapter 2 focuses on the forecasting of time series models using a variety of sample sizes and forecast periods to determine an "optimal" ARIMA model. The subject of chapter 3 will be to develop a selling strategy for commodity futures. This strategy will be incorporated with the "optimal" ARIMA model and the equivalent neural network model to compare their trading performances. Chapter 4 is a summary and conclusion of the entire research. The thesis ends with a reference section and an appendix.

CHAPTER 2

FORECASTING TIME SERIES MODELS USING VARIOUS SAMPLE SIZES AND FORECAST PERIODS

Introduction

It is generally agreed amongst most researchers that the forecasting performance of a time series model begins to deteriorate the further into the future it attempts to forecast. In addition, it is believed that increasing the sample size should enhance the ability of a model to forecast accurately. It is in the interest of researchers to determine the optimal combination of sample size and forecast period which will yield the minimum forecast error. What exactly is this optimal set? It may be tempting to conclude that obviously the largest sample size possible together with the minimum forecast period will produce minimum error. It is only through conducting an experiment with real data involving various combinations of sample size and forecast period that we can begin to understand the relationship between these variables and the forecast error.

The objective of this chapter is to forecast time series models using a variety of different sample sizes and forecast periods. The prime focus will be to see if either of these variables affect the value of the forecast error. Although the general beliefs regarding these relationships may be true it will be interesting to perform a more detailed examination. The results are expected to be in agreement with the hypothesis that larger sample sizes and shorter forecast periods yield smaller forecast errors.

An optimal model based on the minimum forecast error will then be selected. This model will be compared to a neural network model utilizing the same sample size and forecast ahead period. The models will be compared by investigating the changes in the forecast error along with their relative ability to forecast turning points in the market.

Data

Weekly data on Cattle (1976-1995) and Wheat (1976-1995) nearby futures prices obtained from the vendor Tehnical Tools Data Services are used. The data are initially provided in daily form but is converted to weekly to smooth the series and to reduce the amount of computing time necessary. In addition, a rollover technique is used in the formation of the data series. For example if the data represents the December futures price it is converted to the next available contract month as soon as the 20th of November is reached. This is implemented to remove the often volatile movement of price series during the contract month and also because most traders get out of futures obligations at least one month before the expiration date. The conversion of the data created 1036 observations over a twenty year time period.

Procedure

The procedure involves the development of an autoregressive (AR) model. The (AR) model is a good estimate of the ARIMA model as long as the moving average (MA) process is invertible. Since (AR) models are simple to estimate, have well-developed model selection criteria and require limited pretesting, they are the form of ARIMA used here for estimation (Kohzadi and Boyd, 1995).

The optimal lag length is chosen using the Akaike information criteria (AIC) (Akaike, 1981). The AIC is used to determine the lag length because it has the desirable feature of weighing the precision of estimate in relation to parsimony in parameterization of a statistical model (Judge, et al., 1988). The AIC results for each of the two series indicate that the minimum AIC for cattle and wheat occurs at lag 1. In addition, this research will also employ a 6 lag model for each commodity which is arbritrarily chosen. This is implemented since it is felt there is a possibility that a one lag model will not adequately capture the behavior of the time series. Therefore the results presented here will be derived from both a 6 lag and 1 lag model.

The data series for both commodities are initially checked for stationarity using the Dickey-Fuller unit root test (Dickey and Fuller, 1981). The results for both of the original series indicated non-stationarity in the data. The series is then differenced and the unit root test is performed again. The differenced series for both commodities according to the unit root test is stationary. Therefore the ARIMA models are estimated using the differenced data.

Sliding window procedure

The price series for both commodities is forecasted with a sliding window procedure. This process involves selecting a given sample size and estimating an ARIMA model. The model is then forecasted a certain number of periods ahead commonly referred to as the step ahead interval. The data set is then shifted by the value of the step ahead interval. The model is then re-estimated using the same given sample size and a new set of forecasts is generated. This procedure is continually repeated until the end of the data set is reached.

For example, suppose the sample size chosen is 500 and the decision is to forecast 10 periods ahead. Observatons 1...500 will forecast 501...510. The data set is shifted by 10 periods and then observations 11...510 will forecast 511...520. This procedure is then continually repeated until the end of the data set is reached.

The sliding window procedure is performed on the lag 1 and lag 6 ARIMA models for cattle and wheat. The sample sizes chosen are 200, 350, and 500. The step ahead forecast periods are chosen arbitrarily as 5, 10, 20, and 50. These periods combined with the three different sample sizes produces 12 different models for each of lag 1 cattle, lag 6 cattle, lag 1 wheat, and lag 6 wheat ARIMA models.

Forecast Evaluation Methods

There are two criteria used here to evaluate the forecast accuracy of the particular ARIMA model. The first and most commonly used measure of forecast error is the mean squared error (MSE). It is the average of the squared errors over a given forecast period. The formula is given by:

$$\text{MSE} = \frac{\sum_{t=1}^{n}(P_t - A_t)^2}{n} \tag{2.1}$$

Where P_t is the predicted value, A_t is the actual value and N is the total number of forecasts.

The second criterion used is the mean absolute percentage error (MAPE). It is a measure of the average absolute percentage error made by the forecasts. The formula is given by:

$$\text{MAPE} = \frac{\sum_{t=1}^{n}\left|(P_t - A_t) \div A_t\right|}{n} \tag{2.2}$$

Factorial Design

In many instances a researcher is interested in determining the variables or factors which influence a particular response variable. In particular the investigator would like to examine the effects and interactions of many variables simultaneously on a dependent variable. (Anderson and McLean, 1974). In a factorial design all possible levels of the factors are investigated. The analysis of the factorial design is to determine if changes in the levels of a certain factor influence the response variable. Using the above ARIMA models the sample size and the forecast period are the factors and the forecast error is the response variable. In most cases a factorial design is replicated several times to allow for measurement of the random variation. This particular research does not require replication due to the fact that replication of the full factorial design would not produce new observations for any step ahead forecast period. In this case there are two factors and only one observation per cell (ie. no replication) so the linear statistical model is given by (Montgomery, 1991):

$$Y_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ij} \tag{2.3}$$

where Y_{ij} is the response variable, μ is the overall mean effect, τ_i is the effect of the ith level of the row factor A, β_j is the effect of the jth level of column factor B, $(\tau\beta_{ij})$ is the effect of the interaction between τ_i and β_j, and ε_{ij} is a random error component.

For the purposes of this study Y_{ij} is the MSE, τ_i is the effect of the sample size, β_j is the effect of the forecast period, $(\tau\beta_{ij})$ is the interaction term and ε_{ij} is the random error term.

The interaction term if present, can not be directly separated from the error term for estimation. A test for the presence of interaction developed by Tukey (1949a) is possible. The test uses the regression equation:

$$(\tau\beta)_{ij} = \gamma\tau_i\beta_j \tag{2.4}$$

where γ is an unknown constant. If the computed F-statistic is greater than the tabled F-value then the hypothesis of no interaction must be rejected.

An analysis of variance table is then created which demonstrates the importance of the factors on the value of the response variable as well as the possible presence of interaction. Large values of the F-statistics indicate that there is an effect on the value of the MSE with a change in the level of the particular factor. The details of these calculations are provided in the appendix at the end of Chapter 4.

Forecasting turning points

In addition to a model being able to forecast accurately over a given time period, a model should also be able to forecast price direction. A model can be very accurate in terms of forecasting error but may perform poorly when attempting to predict the up and down movements of a futures price series. A common measure to determine if a model is predicting market direction is the ratio of accurate forecasts (RAF). It is simply the proportion of forecasts that are made which correctly predict the direction of the price change. Clearly the higher the ratio the better the model is performing.

A much more rigorous approach to test the ability of a model to forecast market direction is provided by Cumby and Modest (1993). This approach uses a regression equation with the forecast value being the dependent variable and the actual value being the independent variable. It is outlined by Kohzadi et al (1996) and (Ntungo, 1996). Both variables are converted to values of 1 and 0 based on the following rules:

$$A_t = 1 \text{ if } \Delta A_t > 0 \text{ and } A_t = 0 \text{ if } \Delta A_t \leq 0 \tag{2.5}$$

$$F_t = 1 \text{ if } \Delta P_t > 0 \text{ and } F_t = 0 \text{ if } \Delta P_t \leq 0 \tag{2.6}$$

where ΔA_t is the amount of change in the actual variable between time t-1 and t and ΔP_t is the amount of change in the forecast variable for the same period.

The regression equation is then given by:

$$F_t = \alpha_0 + \alpha_1 A_t + \varepsilon_t \tag{2.7}$$

where F_t is the forecasted price direction binary variable and A_t is the actual price binary variable. The regression coefficient α_1 is the slope of the fitted line. Values of the coefficient that are greater than zero indicate that the model is forecasting the market direction correctly. If the coefficient is significantly different from zero then the model is forecasting correctly with a high degree of probability (Ntungo, 1996).

An alternative approach to test the ability of a model to forecast turning points is provided by Cumby and Modest (1993). Their method involves the creation of a two-by-two contingency table:

		Actual Returns	
		$R^*(t) \geq R(t)$	$R^*(t) < R(t)$
Predicted	$R^*(t) \geq R(t)$	n_1	N_2-n_2
Returns	$R^*(t) < R(t)$	N_1-n_1	n_2
	Totals	N_1	N_2

Let $R^*(t)$ denote the return on the first investment over a holding period beginning at time t, and let $R(t)$ denote the return on a second investment. In addition, let N_1 be the number of outcomes with $R^* \geq R$, N_2 be the number of outcomes with $R^* < R$, n_1 be the number of correct forecasts when $R^* \geq R$, and n_2 be the number of correct forecasts of $R^* < R$. Clearly a model is forecasting market direction correctly when n_1 is a high proportion of N_1 and n_2 is a high proportion of N_2.

The contingency table can be analyzed using an approach that is outlined in McClave and Dietrich (1988). The details of the calculations to derive the χ^2 statistic are provided in appendix 2 at the end of the thesis. The test is to determine whether there is a dependent relationship between two variables. A high value of the χ^2 test statistic indicates evidence that there is a dependent relationship between the variables. This research demonstrates whether the values of the actual and forecast variables are moving in unison. In essence this implies that forecasts which predict upward movements produce a "high" percentage of corresponding upward movements in the actual price series. Clearly there is a similar argument for the downward forecasts. It must also be noted that this test can be misleading. A significant χ^2 variable could just as easily signify that the forecast and actual direction are moving in an opposite fashion. A careful examination of the data should be sufficient to verify the interpretation of the χ^2 test statistic. If they are acting independently then a set of forecasts which predict either an upward or downward movement in the price should correspond to a roughly equal amount of upward and downward movements in the actual price series. This means if we forecast a price series to say move upwards 300 times then there will be approximately 150 upward movements and 150 downward movements in the corresponding actual price series. A low value for the test statistic demonstrates that there is not a strong relationship between the forecast direction and the corresponding actual price direction.

Testing RAF Differences

A suitable approach to test whether proportions are significantly different is provided by McClave and Dietrich (1988). This analysis would like to determine if there are differences among the RAF values. In particular we will compare the RAF values for the neural network against its corresponding AR models. In this case there will be four separate tests performed which will compare each neural network against the two other ARIMA models.

The assumption will be that the sampling distribution of the differences in the RAF values is approximately normal since the sample size used here is large (n>30). Therefore we can use the z-statistic to derive confidence intervals and to test the hypothesis that the RAF values are equal.

In general we can conclude there is evidence to indicate a difference among proportions if we can reject the null hypothesis:

$$H_0: (p_1 - p_2) = 0$$

H_1: $(p_1-p_2)<0$ or $(p_1-p_2>0)$

The test statistic is given by:

$$z = \frac{(p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \qquad (2.8)$$

where p_1 and p_2 are the proportions of correct decisions and $q_1=1-p_1$ and $q_2=1-p_2$. The sample sizes are n_1 and n_2. The denominator is simply the standard deviation of the difference in proportions. Absolute values of the test statistic which are greater than 1.645 indicate there is a significant difference in the proportions at $\alpha=0.05$.

Results

Price forecast performance

Table 2.1 shows results for the mean square error of the ARIMA lag 1 model for wheat futures. It is evident from the data that an increase in the forecast period increases the mean square error. Each sample size of 500, 350, and 200 showed significant enlargement in the mean square error as the forecast ahead period was increased. This supports the hypothesis that the forecast error of a model increases the further ahead it attempts to forecast. The minimum MSE of 291.1 was achieved using a sample size of 350 and a forecast period of 5..

Table 2.1 also provides evidence that increasing the sample size may not be an important criterion when attempting to reduce the mean square error. Notice specifically that as the sample size was increased from 350 to 500 that there was actually an increase in the MSE. Interestingly there was a reduction in the MSE as the sample size was increased from 200 to 350. This indicates that a certain sample size may be necessary for accurate forecasting but increasing an already large sample size provides no benefits in reducing the forecast error.

Table 2.2 shows results for the MSE of the ARIMA lag 6 model for wheat futures. The results are extremely similar to those found in table 2.1. This indicates that a lag 6 model does not perform any better than the lag 1 model. In addition, the results indicate on average that increasing the forecast period increases the MSE and that an increase in the sample size does not guarantee a decrease in the MSE. The minimum MSE of 276.8 was obtained using a sample size of 350 and a forecast period of 5 which is the same combination as the lag 1 model.

Tables 2.3 and 2.4 show the equivalent results for cattle futures. The values of the MSE shown in these tables is significantly less than the results computed for the wheat futures. This is due to the fact that wheat futures data are much larger in magnitude and also more volatile. It is clear from these results that there is virtually no difference in the performance of a 6 lag model versus a 1 lag model. Similarly it can be observed that in general an increase in the forecast period increases the value of the MSE. It is also evident that increasing the sample size produces mixed results for the MSE. In some cases increasing the sample size reduced the MSE while in other cases it increased the MSE. Note that the minimum MSE was again obtained for both the 6 and 1 lag model using a sample size of 350 and a forecast period of 5.

Table 2.5 shows wheat futures results of the MAPE for the ARIMA lag 1 model. The results here are very similar to table 2.1. The MAPE increases as the forecast period increases. The value of the MAPE remains somewhat constant across various sample sizes. Note in particular forecast period 5 which produced values of 3.61%, 3.73% and 3.88% for sample sizes of 200, 350, and 500 respectively. The minimum MAPE of 3.61% occurred for a sample size of 200 with a forecast period of 5.

Table 2.6 shows the MAPE results for the wheat futures ARIMA lag 6 model. It appears again that the 6 lag performs equally as well as the 1 lag model. In addition, there is an apparent upward trend in the results as the forecast period increases. The sample size again produces mixed results with up and down movements of the MAPE occurring as the sample size increases. The minimum MAPE of 3.62% occurred using a sample size of 350 and a forecast period of 5.

Tables 2.7 and 2.8 show the MAPE results for the cattle futures data. In overall terms the MAPE for cattle is less than that of wheat which again is due to the high degree of volatility in the wheat futures data. The general trend is that decreasing the forecast period will decrease the MAPE. The results from this table give some evidence that increasing the sample size results in a decrease in the forecast error. Notice in table 2.7 for the 50 forecast ahead period that the MAPE decreases as you move from sample size 200 to 500. In addition table 2.8 which outlines the lag 6 model indicates a somewhat general downward trend in the MAPE as the sample size increases.

Comparing Neural Network Versus Arima Model

Table 2.9 Provides the forecast error results for the two arima models and the neural network model for both the cattle and wheat commodities. Notice that the models chosen employ a sample size of 350 and a step ahead interval of five. This was the optimal combination that was chosen due to the fact that it had the minimum forecast error among all the developed ARIMA models.

It is readily apparent from these results that the neural network outperforms the ARIMA models. When comparing it against the ARIMA lag 6 model it reduces the MSE from 276.8 to 120.3 for the wheat futures data. This is in agreement with Kohzadi and Boyd, (1995) who found decreases in the MSE using weekly corn futures data. These results are also similar to Dematos (1996) who found lower MSE values using a neural network against an ARIMA model using monthly Japanese yen futures prices. In addition, it lowered the MAPE from 3.62% to 2.43%. There were also large gains in the forecast error for the cattle futures data. The MSE is lowered to 2.773 compared to 6.897 and 6.915 for the ARIMA lag 1 and lag 6 models respectively. It is also able to lower the MAPE to 1.92% from 2.94% that is computed from both the lag 1 and lag 6 ARIMA models.

Factorial Design Results

The rest of the tables for this section can be found in appendix A.3. The appendix includes tables 2.10 to 2.17 inclusive.

This section outlines the analysis of variance tables for the test on the effects of forecast periods and sample size on MSE which are shown in tables 2.10-2.13. A significant F-statistic (i.e. a value that is larger than the value in the F-statistic tables at a 5% level), indicates that the factor is making an important contribution in determining the MSE. This means for example that an increase in the value of the variable will produce a significant change in the value of the MSE. This is not to say that the MSE will also increase. It could very well imply that the MSE is moving downward with changes in the predictor variable. The essence is that the change in the MSE will be significant. Small values of the F-statistic indicate that the variable is not significant in determining the value of the MSE. This implies that changes in the predictor variable whether they are increases or decreases will not significantly alter the value of the MSE.

Table 2.10 is the analysis of variance table of the lag 1 wheat futures model. This table demonstrates the relative importance of each factor on the response variable MSE. The forecast interval is a very significant variable as it has a computed F-statistic of 227.2. This is much greater than the tabled value of $F_{3,11,.05} = 3.59$. The sample size variable with a computed

F-statistic of 3.17 is less than the tabled value of $F_{2,11,.05} = 3.98$. Therefore the effect of the sample size on the MSE is not significant. This is evident from tables 2.1 to 2.4 data which indicates a fairly constant MSE across the various chosen sample sizes.

Table 2.11 is the analysis of variance table of the lag 6 wheat futures model. The results here are very similar to table 2.9 as the forecast interval is again significant at the 5% level while the sample size variable is not significant.

Table 2.12 shows the analysis of variance for the lag 1 model of the cattle futures data. The computed F-statistic of 11.17 is significant at the 5% level. The computed F of 1.36 for the sample size variable is not significant at the 5% level.

Table 2.13 is the analysis of variance for the lag model of the cattle futures data. Once again the forecast interval is shown to be a significant variable while the sample size is not significant at the 5% level.

Tables 2.14 and 2.15 summarize the results of the Merton's Test to indicate the ability of a model to forecast market direction. Note that only the Arima Lag 1 model for wheat has a positive value for the coefficient of α_1 but it was not a significant variable. This implies that it correctly forecasted market direction but with low probability. The other interesting results are derived from the Cattle Lag 1 model. It has a negative coefficient for α_1 and it is significant at the 5% level. This means that it significantly forecasted market direction incorrectly.

Table 2.14 also includes the RAF values for the wheat futures among the three models. The neural network actually performs the worst among the three models as it was only able to forecast the market direction correctly 48.82% of the time. The lag 1 and lag 6 models with values of 51.09% and 49.20% performed slightly better.

Table 2.14 similarly provides the results for cattle futures. In this scenario the neural network with an RAF of 49.78% performed slightly better than the lag 1 and lag 6 models which produces RAF values of 47.15% and 47.59% respectively.

The RAF values computed here are similar to Ntungo (1996). His study utilized corn, silver, and Deutsche Mark weekly futures prices and compared 10 and 25 step ahead forecast periods using ARIMA models and a neural network.

Table 2.16 is a summary of the results for the test of independence between the forecast and actual price variables. The lag 1 model for cattle had a significant χ^2 value of 4.845. This was mainly due to the fact that this model was able to correctly forecast 66% of the upward movements in the price series. The remaining test statistic values were not significant at the 5% level. This is not surprising due to the fact that the models are in general only forecasting the market direction correctly about 50% of the time as is evidenced by the RAF values

The last set of results are given in table 2.17 which outline the test for the differences of the ratio of accurate forecasts. The four sets of tests are computed which compare the neural network model for each commodity against its corresponding lag 1 and lag 6 ARIMA models. The values of the test statistics for all four combinations are not significant which indicates that the neural network does not significantly outperform the ARIMA models. This is not a surprising result due to the fact that the RAF values for these models were very similar.

Conclusions

The purpose of this chapter is to investigate the relative importance of the sample size and the forecast ahead interval in determining the forecast error of time series models. In addition, the optimal AR model was chosen and compared with a neural network model. Nearby weekly futures closing prices for cattle and wheat are used. The ARIMA models are developed by using the Akaike information criteria which determines the optimal lag length. In this case the minimum AIC value occurs with a lag length of one. The study also incorporates a lag length of six which is arbritrarily chosen to allow for the fact that a one lag model may be inadequate for forecasting purposes as it only considers prices that are one period back and thus may be missing important information. ARIMA models are estimated employing five, ten, twenty and fifty forecast ahead periods together with sample sizes of 200, 350, and 500. This produces twelve different values for the forecast error for each of the lag 1 cattle, lag 6 cattle, lag 1 wheat, and lag 6 wheat models.

The general result apparently indicates that the forecast error increases as the forecast ahead period increases. This agrees with the hypothesis that forecast accuracy is compromised the further ahead the forecast is. The sample size appears to not be an important factor in determining the forecast error. Increasing the sample size does not lower the forecast error as expected. In some cases it actually increases the forecast error. The lag 1 and lag 6 cattle models demonstrate a decrease in the forecast error as sample size increases but only when using a forecast ahead period of 50. This indicates that increasing the sample size may be helpful for longer range forecasts.

The neural network model showed a vast improvement over the comparable ARIMA models as it was able to reduce the forecast error. Both the cattle and wheat models demonstrate a large reduction in both the MSE and the MAPE. This strongly supports the hypothesis that the neural network should be able to produce lower forecast errors than the ARIMA models.

The ARIMA models and the neural network models appear to perform equally well when attempting to forecast the market direction. The RAF measure turns out to be roughly the same for both the neural network and the ARIMA models. The neural network is slightly better at testing market direction for the cattle futures but does not perform as well as the ARIMA models for the wheat futures data.

The Merton's test seems to indicate that neither the ARIMA model or the neural network are able to forecast market direction correctly. In fact the cattle lag 1 model predicts incorrectly with high probability.

The test of independence shows that only the cattle lag 1 model was doing an adequate job of predicting market direction. This although was largely due to the fact that it predicts a very high percentage of the upward movements in the data. Overall though this model performes only slightly better than the other models. Finally the test to determine if there was a significant difference in the RAF values for the ARIMA models against the neural network came out inconclusive. The neural network does not perform significantly better than the ARIMA models. This does not support the hypothesis that a neural network should be able to forecast market direction more accurately.

Table 2.1 Mean square error results of ARIMA model lag 1 with varying sample sizes and Forecast Periods, Weekly Wheat Futures Closing Prices, Chicago Mercantile Exchange, 1976-1995

	Sample Size		
Forecast Ahead Period (Weeks)	500	350	200
50	3453.2	3168.9	3629.3
20	1541.0	875.4	1358.6
10	542.4	480.1	519.4
5	306.9	291.1	296.4

Table 2.2 Mean square error results of ARIMA Model lag 6 with varying Sample Sizes and Forecast Periods, Weekly Wheat Futures Closing Prices, Chicago Mercantile Exchange, 1976-1995

	Sample Size		
Forecast Ahead Period (Weeks)	500	350	200
50	3422.1	3189.3	3835.7
20	1547.5	890.6	1375.1
10	550.5	494.0	537.9
5	307.3	276.8	344.5

Table 2.3 Mean square error results of ARIMA model lag 1 with varying Sample Sizes and Forecast Periods, Weekly Cattle Futures Closing Prices, Chicago Board of Trade, 1976-1995

	Sample Size		
Forecast Ahead Period (Weeks)	500	350	200
50	14.550	17.507	25.222
20	15.811	21.834	18.126
10	13.057	12.146	12.511
5	7.268	6.897	8.117

Table 2.4 Mean square error results of ARIMA Model Lag 6 with varying Sample Sizes and Forecast Periods, Weekly Cattle Futures Closing Prices, Chicago Board of Trade, 1976-1995

	Sample Size		
Forecast Ahead Period (Weeks)	500	350	200
50	14.566	17.350	23.986
20	15.804	20.154	18.584
10	13.229	12.345	12.787
5	7.204	6.915	7.912

Table 2.5 Mean absolute percentage error results of ARIMA Model Lag 1 with varying Sample Sizes and Forecast Periods, Weekly Wheat Futures Closing Prices, Chicago Mercantile Exchange, 1976-1995

	Sample Size		
Forecast Ahead Period (Weeks)	500	350	200
	%	%	%
50	12.12	11.78	12.64
20	8.75	6.70	7.89
10	5.23	4.83	4.90
5	3.88	3.73	3.61

Table 2.6 Mean absolute percentage error results of ARIMA Model Lag 6 with varying Sample Sizes and Forecast Periods, Weekly Wheat Futures Closing Prices, Chicago Mercantile Exchange, 1976-1995

	Sample Size		
Forecast Ahead Period (Weeks)	500	350	200
	%	%	%
50	12.08	11.81	12.85
20	8.76	6.77	7.94
10	5.27	4.90	4.98
5	3.86	3.62	3.97

Table 2.7 Mean absolute percentage error results of ARIMA Model Lag 1 with varying Sample Sizes and Forecast Periods, Weekly Cattle Futures Closing Prices, Chicago Board of Trade, 1976-1995

	Sample Size		
Forecast Ahead Period (Weeks)	500	350	200
	%	%	%
50	4.41	4.86	5.87
20	4.40	5.42	4.88
10	4.01	3.93	4.06
5	2.93	2.94	3.19

Table 2.8 **Mean absolute percentage error results of ARIMA Model Lag 6 with varying Sample Sizes and Forecast Periods, Weekly Cattle Futures Closing Prices, Chicago Board of Trade, 1976-1995**

	Sample Size		
Forecast Ahead Period (Weeks)	500	350	200
	%	%	%
50	4.37	4.81	5.72
20	4.40	5.19	4.93
10	4.03	4.00	4.12
5	2.91	2.94	3.17

Table 2.9 Forecast Error Results for the ARIMA Models and a Neural Network Model, Weekly Futures Closing Prices, 1976-1995

WHEAT 1975-1995

Model	MSE	MAPE(%)	Sample Size	Steps Ahead
Arima Lag 1	291.1	3.73	350	5
Arima Lag 6	276.8	3.62	350	5
Neural Network	120.3	2.43	350	5

CATTLE 1975-1995

Model	MSE	MAPE(%)	Sample Size	Steps Ahead
Arima Lag 1	6.897	2.94	350	5
Arima Lag 6	6.915	2.94	350	5
Neural Network	2.773	1.92	350	5

CHAPTER 3

COMPARING TRADING PERFORMANCES OF AN ARIMA MODEL AND A NEURAL NETWORK MODEL

Introduction

There has been a fair bit of research done over the last several years which attempt to compare the forecasting accuracy of various forecasting techniques (Mahmoud, 1984). For example (Brandt and Bessler, 1984) demonstrated that ARIMA models are better at predicting price changes as opposed to vector auto regressive models. Another study by (Dorfman and McIntosh, 1990) compared a variety of ARIMA models and vector autoregressive models and found that no method dominated the other in terms of forecasting accuracy.

Recently there has been an increasing interest in using neural networks for forecasting. They utilize a non-linear approach to forecasting and are particularly well suited to modelling futures data which often displays chaotic behaviour. For example Dematos et al (1996) shows neural networks outperform ARIMA models when forecasting Japanese Yen futures. A study by (Grudnitski and Osburn, 1993) determined that neural networks are particularly well suited to finding accurate solutions in an environment characterized by complex, noisy, irrelevant or partial information. In addition, (Kohzadi et al, 1996) found that neural network out perform ARIMA at forecasting market turning points.

Since neural networks have been able to outperform ARIMA models in terms of forecasting it would be interesting to see how they perform against traditional forecasting techniques using a simple trading model. Clearly a model developed for futures price prediction is only useful if it can trade successfully in the market. A variety of studies over the years have utilized various econometric and time series models to trade on the futures markets. One such study by (Kastens and Schroeder, 1995) utilizes a simple trading rule and a basic regression model to produce positive profits while trading on the cattle futures market.

A number of studies were performed to determine the ability of a neural network to trade commodity futures. One such study by (Hamm et al, 1993) shows that 3 of the 5 trading models produced statistically significant returns using a neural network. Another study by (Mendelsohn and Stein, 1991) trains a neural network on 3 years of daily D-Mark futures prices to generate significant profits net of transaction costs.

The purpose of this study is to compare the trading results of an ARIMA model against a neural network model. Specifically the study will explore whether either model can produce significant profits per trade. The study will also check to see if neural network models trade more often than ARIMA model using a given trading strategy. In addition the study will try to determine if either the ARIMA model or the neural network model can sell at a significantly higher price than the overall price utilizing the given trading regime. The chapter begins with a brief introduction into the process of creating a neural network model. Following this is an explanation of the trading strategy employed followed by a conclusion which sumarizes the results. It is expected that the neural network will perform more trades over the given time interval due to the fact that it should be able to react to market direction changes in a quicker fashion. In addition, it is hypothesized that neural networks should produce higher profits than the ARIMA model due to its ability to capture the behaviour of non-linear data which is typical of futures price series. Thirdly it is anticipated that neural networks will sell at significantly higher prices than the overall selling price.

Preparation of the Neural Network

A paper done by Kaastra and Boyd (1996) outlines a design procedure to develop a commodity futures neural network forecasting model.

Input Selection

Each commodity in this study has a separate neural network developed for its specific forecasting purposes. The inputs to each neural network are the number of lags which are determined from the Akaike Information Criterion (AIC). A six lag model for cattle and wheat was arbitrarily chosen as a suitable lag value. The AIC criterion indicated a one lag model that severely limits the development of the neural network model.

Data Collection and Preparation

Weekly data on cattle in cents per pound (1976-1995) and wheat in cents per bushel (1976-1995) for nearby futures closing prices are obtained from the vendor Technical Tools Data. The first 350 observations or about 7 years of data are used to estimate the model. The data spans 1036 observations which leaves 686 to be used for forecasting purposes. The model is continually re-estimated every 5 weeks to generate 5 forecasts at a time until the end of the data set is reached.

The data was converted to values between 0 and 1 based on a minimum/maximum rule. The largest value was given the value of 1 while the smallest value was scaled to zero. The rest of the data points in between were proportionally mapped between 0 and 1. This formatting is done to better allow the network to memorize the patterns of the data.

Contract months are rolled over approximately a month before expiration to avoid noisy data which is somewhat typical of futures prices that are less than a month from expiration. The optimal lag length was determined using the Akaike Information Criterion (AIC). A one lag model was determined for both the Cattle and Wheat data. For the purposes of this study a six lag model will be used. This is a purely arbitrary choice but should be adequate for modelling purposes as it considers prices that are six periods back. A one lag model may be inappropriate due to the fact that it does not consider prices that are more than one period back. The unit root test which is used to test for stationarity indicated that the first differenced series was stationary. Therefore the differenced series will be used for the analysis.

Training and Testing Sets

The neural network model is estimated using 350 observations of weekly data in sample. This is rougly equivalent to seven years of data. The neural network training uses 90 percent of the 350 observations and the remaining 10 percent are used for neural network testing.

Neural Network Design.

The most common neural network is the three layered back-propagation neural network and it is the form used here for each commodity. The cattle neural network has six input neurons representing the six lags. It has one hidden layer with six neurons and one output layer with one neuron. The wheat neural network has six input neurons, one hidden layer with six neurons and one output layer with one neuron.

Neural Network Evaluation Criteria

Several neural networks are created for each commodity. The difference in each neural network architecture is the number of neurons within each hidden layer. There are six possible neural networks for each commodity since there are six input neurons for each commodity. The restriction is that the number of neurons within the hidden layer can not exceed the number of input neurons. Evaluating the different neural networks involves looking at the mean square error. The optimal neural network that is chosen is the one with the minimum mean square error. The mean square error values are obtained by passing the data set through the neural network for a total of 500 runs which is a default setting in the program. The theory is that as the neural network continues to make passes through the data set it begins to recognize patterns in the data and thus enables it to continually lower the mean square error.

Neural Network Training

The training is automated using the built in features of N-Train Version 1.0. An automated sliding window training technique is employed (Kohzadi, et al. 1996; Kaastra and Boyd, 1995) for each neural network. This sliding window technique is discussed in more detail in the procedure section of chapter 2.

Implementation of the Trained Neural Network

Using the neural network involves comparing its forecasting results with out of sample data. The forecasting is done using a five step ahead interval. The forecasts and the corresponding actual values are used later for analysis of the trading system.

ARIMA Model Development and Forecasting Procedure

The details of the development of the ARIMA model are provided in the procedure section of chapter 2. Recall that the sample size of 350 with a step ahead interval of five were considered to be the optimal models based on their minimum mean square errors. This model coupled with the neural network model using the same sample size and step ahead interval will be used to perform a comparative analysis of the trading results.

Trading Model

Constructing a Valid Trading Simulation

A text by (Schwager, 1984) outlines several important issues that need to be considered when constructing a trading model.

1. The longer the time periods of the trading signals the longer the data set needs to be.
2. A trading rule should be as simple as possible. The more rules a system has, or the more conditions that have to be met for a supposed trade to occur, the less likely it is that an identical situation will occur in the future.
3. The profits generated from the trading need to be sufficient to cover the costs of transcations, system design etc.

The trading system in this research will hopefully meet the 3 different criteria. The data set of 686 observations should be large enough to generate many trading signals. The trading rule is straightforward and will be discussed further in the next section. The system will hopefully be able to generate profits that can cover all of the necessary costs.

Trading Model Assumptions

There are two assumptions that are in effect for this trading system. Firstly, the system will only buy or sell one contract at a time. Secondly the nearby futures contract is always the one being traded.

Trading Strategy and Rule

There are a number of strategies that can be employed to trade on the futures market. The approach used here is a straight forward approach which uses the forecasted prices, P_{t+1}, as a signal to buy and sell futures contracts. The rule for trading is outlined as follows:

If $P_{t+1} < P_t$ then sell.

where P_{t+1} is the forecasted weekly futures closing price for time t+1 and P_t is the forecasted weekly futures closing price for time t. This position is maintained in the market until $P_{t+1} > P_t$. At this point the futures contract is bought back and the trader exits the market. The profit or loss is calculated and the trader waits for the next available sell signal. The trader then sells the contract and waits again until the price forecast is for it to rise. The market is then exited by buying back the contract. This process is continually repeated until the end of the data set is reached.

The returns on each trade are calculated from net contract values which are determined by subtracting transaction costs that are made up of two parts. A fifty dollar transaction fee for each buy and sell is subtracted from each profit amount. Secondly both wheat and cattle have seven trading months during the year. A one hundred dollar transaction fee is also included to allow for the fact that the contract needs to be rolled over into the next available trading month approximately one month before the contract expiration date.

Trading Statistics and Significance Tests

The trading statistics are collected over a 686 week period for both the cattle and wheat futures data. The statistics include the total number of trades, the number of profitable trades, the number of losing trades, the number of profitable trades, the average profit per trade, the average profit per profitable trade, the average loss per losing trade. In addition the percentage of profitable and losing trades are also included.

There are two tests which involve employing the two-tailed t-test. The first is checking whether the average profit per trade is significantly different from zero. The t-test should be suitable since the sample size used here is sufficiently large. The formal test involves declaring a null hypothesis versus an alternative hypothesis. The test outlines as follows and is similar to (Ntungo, 1996):

H0: APPT=0

H1: APPT≠0

where APPT represents the average profit per trade for a given commodity. This hypothesis can be tested using the student's t-statistic. The general form of the t-statistic is given by:

$$t = \frac{\bar{x} - x_0}{s/\sqrt{n}} \tag{3.1}$$

where $\bar{x}$ is the average profit per trade, x_0 is the expected profit per trade, which is assumed to be zero under HO, s is the sample standard deviation, and n is the sample size. Large values of the t-statistic which are greater than the corresponding tabled t-value indicate that the profits generated are significantly greater than zero. Similarly large negative values of the t-statistic which are less than the corresponding negative t-value from the table indicate that profits are significantly less than zero.

The second test is to check if the difference in the selling price obtained from the trades is different from the actual selling price over that particular interval. This can be formally tested by using the null hypothesis:

H_o: ASPTM-ASPO =0

H_1: ASPTM-ASPO $\neq$0

where ASPTM is the average traded selling price over the particular time period using the trading model, and ASPO is the average actual selling price overall over the particular time period. The t-statistic is calculated in a similar manner as above where large values of the t-statistic indicate that the trading model is selling at prices that are significantly higher than the average selling price. Similarly a significant negative value for the t-statistics indicates the trading model is selling at prices which are less than the average price.

Results

Trading Results

Table 3.1 summarizes the trading statistics for the cattle futures data. The neural network did initiate more trading signals as it traded 177 times as opposed to the ARIMA model which traded 160 times. Both models were unable to generate positive profits as indicated by the average $209.45 average loss per trade for the ARIMA model and an average $150.62 loss for the neural network. The neural network did at least improve the average profit but was unable to generate positive profits. The ARIMA model was slightly better at producing profitable trades than the neural network but its success rate of 41.88% was quite low. Is is also interesting to note that the neural network although unable to produce as high a frequency of profitable trades, it was able to perform slightly better on average when performing a profitable trade. This is indicated by its average profit per profitable trade of $572.59 as opposed to the ARIMA model of $506.70. These results are further enhanced by the fact that the neural network's average loss on losing trades was $623.75 which was less than the ARIMA model which had an average loss of $725.38.

Table 3.2 shows the trading statistics for the wheat futures data. The neural network traded only slightly more often than the ARIMA model. The average profits for both the ARIMA model and the neural network were again negative. Although again the neural network's loss was less indicated by an average $168.89 loss per trade as opposed to the ARIMA model with an average loss of $238.70 per trade. Once again the neural network performed poorly compared to the ARIMA model as it produced only positive profits 36.67%

of the time compared to the ARIMA model with a value of 42.61%. The results were also similar to the cattle future results. The neural network produced higher profits per trade for the profitable trades and smaller losses on average for the losing trades.

<u>Tests of Significance Results</u>

Table 3.3 is a summary of the results which shows whether the aveage profit per trade for a given model is significantly different from zero. The cattle futures results have both the ARIMA model and the neural network model showing significant losses. This is evidence by the t-statistics which are less than -1.96 which is the critical value.

Table 3.4 shows the results for the wheat futures. It is evident again that both the neural network model and the ARIMA model have negative profits which are significant. Both values of the t-statistics are less than -1.96.

Table 3.5 shows the significance test to see whether the average selling price using the trading model is significantly different than the average selling price overall. It is evident from this table that the trading model was unable to sell at a significantly higher price than the overall average selling price. The ARIMA model did sell at a higher price but was not significantly higher since the t-statistic was 0.819 which is less than the critical value of 1.96. The neural network model actually traded at a lower average price than the overall price but it was not significantly less as evidenced by the t-statistic value of -0.071.

Table 3.6 summarizes the results for the wheat futures. Both the ARIMA model and the neural network model traded on average at a lower selling price than the overall selling price. Both t-statistics were again not significant at the 5% level.

Summary

The objectives of this study are to compare the trading performances of an ARIMA model and a neural network model. The second objective is to determine if the neural network model trades more often than than the ARIMA model. Thirdly this study focuses on whether an ARIMA model or a neural network model can sell at a significantly higher price using a given trading model. In essence, we want to see if the models can pick out the high points in the market.

This study shows three important findings. The neural network produces higher profits per trade than the ARIMA model. This agrees with Ntungo (1996) who found higher profits for silver futures when trading using a neural network as opposed to an ARIMA model. His study also investigated corn and Deutsche Mark futures and he found the neural network was unable to outperform a similar ARIMA model. His findings may be different than the results here due to the fact that he used different step ahead forecast intervals and a different trading strategy.

It is interesting to note that in spite of the fact that ARIMA models produce less profits per trade in this study they do predict a higher percentage of profitable trades. This may indicate that the neural network is able to predict large and more important turning points and thus avoid large losses in trades.

Secondly, the neural network trades more often than the ARIMA model using cattle and wheat futures but this difference is very minimal. This does not contradict the hypothesis that a neural network should trade more often due to its ability to react to turning points.

Thirdly this study demonstrates that neither model was able to sell at the high points in the market. Only the ARIMA model for cattle futures sells at a higher average price than the overall average selling price but it is not statistically significant.

Table 3.1 Trading Results of Cattle Futures For an ARIMA model and a Neural Network Model Using a Five Steps Ahead Interval For Forecasting, Weekly Prices, (1976-1995).

Statistic	ARIMA 5	NEURAL NET 5
Total number of trades	160	177
Average profit per trade ($)	(209.45)	(150.62)
Total number of profitable trades	67	70
Percentage of profitable trades	41.88	39.55
Average profit per profitable trade($)	506.70	572.59
Total number of losing trades	93	107
Percentage of losing trades	58.12	60.45
Average loss per losing trade ($)	(725.38)	(623.75)

Table 3.2 Trading Results of Wheat Futures For an ARIMA model and a Neural Network Model Using a Five Steps Ahead Interval For Forecasting, Weekly Prices, (1976-1995).

Statistic	ARIMA 5	NEURAL NET 5
Total number of trades	176	180
Average profit per trade ($)	(238.75)	(168.89)
Total number of profitable trades	75	66
Percentage of profitable trades	42.61	36.67
Average profit per profitable trade($)	339.97	482.96
Total number of losing trades	101	114
Percentage of losing trades	57.39	63.33
Average loss per losing trade ($)	(668.49)	(546.28)

Table 3.3 Tests of Significance for Average Profit Per Trade of Cattle Futures for an ARIMA Model and a Neural Network Model, Weekly Prices, (1976-1995)

Model		
ARIMA 5	Mean	t-statistic
H_o: APPT=0	($209.45)	-3.19 *
H_1: APPT≠0		
Neural Net 5	($150.62)	-2.51 *
H_o:APPT=0		
H_1:APPT≠0		

APPT- Average Profit Per Trade
* Significant at the 5% level
Numbers in brackets indicate negative values

Table 3.4 Tests of Significance for Average Profit Per Trade of Wheat Futures for ARIMA and Neural Network Models, Weekly Prices, (1976-1995)

Model		
ARIMA 5	Mean	t-statistic
H_o: APPT=0	($238.70)	-4.79 *
H_1: APPT≠0		
Neural Net 5	($168.89)	-3.35 *
H_o:APPT=0		
H_1:APPT≠0		

APPT- Average Profit Per Trade
* Significant at the 5% level
Numbers in brackets indicate negative values

Table 3.5 Tests of Significance For The Difference Between Average Selling Price of the Trading Model versus Average Selling Price Overall for Cattle Futures using an ARIMA Model and a Neural Network Model, Weekly Prices, (1976-1995)

Model	Mean Difference	t-statistic
Arima 5 H_o: ASPTM-ASPO=0 H_1: ASPTM-ASPO≠0	0.47	0.819
Neural Net 5 H_o: ASPTM-ASPO=0 H_1: ASPTM-ASPO≠0	-0.04	-0.071

ASPTM- Average Selling Price Using the Trading Model
ASPO- Average Selling Price Overall
All t-statistics are not significant at the 5% level.

Table 3.6 Tests of Significance For The Difference Between Average Selling Price of the Trading Model versus Average Selling Price Overall for Wheat Futures using ARIMA and Neural Network Models, Weekly Prices, (1976-1995)

Model	Mean Difference	t-statistic
Arima 5		
H_o: ASPTM-ASPO=0	-0.75	-0.170
H_1: ASPTM-ASPO≠0		
Neural Net 5		
H_o: ASPTM-ASPO=0	-5.58	-1.335
H_1: ASPTM-ASPO≠0		

ASPTM- Average Selling Price Using the Trading Model
ASPO- Average Selling Price Overall
All t-statistics are not significant at the 5% level.

CHAPTER 4

SUMMARY

The objectives of this study are to (1) to determine if the sample size and the step ahead forecast period are significant in effecting the forecast error; (2) to determine if a neural network model can produce lower forecast error compared to an ARIMA model; and (3) to compare the relative trading performance of an ARIMA model and a neural network model.

Weekly data for cattle in cents per pound (1976-1995) and wheat in cents per bushel (1976-1995) of nearby futures closing prices are obtained from the vendor Technical Tools Data. The price series is created by rolling over the futures price into the next available contract approximately one month before expiration. The weekly series was generated by using the Tuesday price of each week. The first seven years of data are used to estimate the models and leaving out of sample results for cattle and wheat (1983-1995).

Five, ten, twenty, and fifty steps ahead along with sample sizes of 200, 350, and 500 are utilized to choose an optimal ARIMA model. The five step ahead interval in combination with the sample size of 350 produces the lowest forecast error and is used to compare against the neural network. Both the ARIMA model and the neural network model use a lag length of six which is chosen arbitrarily. These models are both used to generate the trading performance results.

Forecasting Time Series Models Using Various Sample Sizes and Forecast Periods

Chapter 2 investigates the relative importance of sample size and forecast period in determining the forecast error. The optimal model is then compared against the neural network model. It turns out that sample size has no significant bearing on the forecast error but it is evident that increasing the forecast period does increase the forecast error. The neural network model is successful in lowering the forecast error compared to the ARIMA model. The neural network model and the ARIMA model forecast turning points correctly at about the same level.

Comparing Trading Performances of an ARIMA model and a Neural Network Model

Chapter 3 provides a brief description of the development of a neural network model. This is followed by an explanation of the trading model. It turns out that the ARIMA and neural network models produce negative average profits per trade for both cattle and wheat futures. The neural network though suffers less losses on average and thus indicates that a neural network can improve trading performance in this regard. The neural network model trades slightly more often than the ARIMA model but the difference is very minimal.

Both the ARIMA and neural network models are unable to sell at significantly higher prices than the overall selling price. This indicates that the trading model developed is possibly not appropriate.

Suggestions for Further Research

The development of the ARIMA model could be improved by utlizing smaller steps ahead intervals accompanied by smaller sample sizes. It is possible that reducing the steps ahead period down to values less than five would continue to produce lower forecast errors. The use of smaller sample sizes may possibly show that sample size is important in determining forecast error. A study utilizing sample sizes between say 20 and 200 may produce very different results.

The trading model in this study is clearly not adequate. The development of a possibly more sophisticated trading rule may generate positive rather than negative profits per trade.

This research could also be expanded by using more commodities. In addition, there are now lots of different neural network programs available and these may produce more conclusive results.

REFERENCES

Akaike, H. "Likelihood of a Model and Information Criteria," Journal of Econometrics 16(1981):3-14.

Anderson, V.L. and McLean R.A. Design of Experiments: A Realistic Approach. Volume 5. New York, 1974.

Blank, Steven C. "Chaos in futures markets? A nonlinear dynamic analysis." The Journal of Futures Markets 11 (1991): 711-728.

Brandt, J.A. and D.A. Bessler. "Forecasting with Vector Autoregressions versus a Univariate ARIMA process: an Empirical Example with US Hog Prices," North Central Journal of Agricultural Economics 6(1984):29-36.

Chavas, Jean-Paul and Matthew T. Holt. "On nonlinear dynamics: The case of the pork cycle." American Journal of Agricultural Economics 73(1991):819-828.

Cumby, R.E. and D.M. Modest. "Testing for Market Timing Ability: A Framework for Forecast Evaluation," Journal of Financial Economics 19(1987): 169-189.

DeMatos G., M. S. Boyd, B. Kermanshahi, N. Kohzadi, and I. Kaastra. "Feedforward Versus Recurrent Neural Networks for Forecasting Monthly Japanese Yen Exchange Rates," Financial Engineering and the Japanese Markets 3(1996):59-75.

Dickey, D.A., and W.A. Fuller. "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," Econometrica 49(1981):1057-1072.

Dorfman, J.H. and C.S. McIntosh. "Results of a Price Forecasting Competition," American Journal of Agricultural Economics 72(1990):804-808.

Grudnitski, G. and L. Osburn. "Forecasting S&P and Gold Futures Prices: An Application of Neural Networks," Journal of Futures Markets 13(1993):631-643.

Hamm, L., B. Wade Brorsen and Sharda, R., "Futures trading with a neural network NCR-134 Conf. on Applied Commodity Analysis, Price Forecasting, and Market Risk Management Proc. Chicago (1993) 286-296.

Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lutkepohl, and T.C. Lee. Introduction to the Theory and Practice of Econometrics. 2nd. ed. New York: John Wiley and Sons, 1988.

Kaastra, I. and M. S. Boyd. "Designing a Neural Network for Forecasting Financial and Economic Timer Series," Neurocomputing 10(1996):215-236.

Kastens and Schroeder. "A Trading Simulation Test For Weak Form Efficiency in Live Cattle Futures." Journal of Futures Markets 15(1995):649-675.

Kohzadi, N., M. S. Boyd, B. Kermanshahi, and I. Kaastra. "Forecasting Livestock Prices with Artificial Neural Networks Versus Linear Time Series Models," Applied Commodity Price Analysis, Forecasting, and Market Risk Management, Conference Proceedings, NCR-134, Chicago Mercantile Exchange, 1994.

Kohzadi, N., M.S. Boyd, I. Kaastra, B.S. Kermanshahi, and D. Scuse. "Neural Networks for Forecasting: An Introduction," Canadian Journal of Agricultural Economics 43(1995):463-474.

Kohzadi, N., M.S. Boyd, B. Kermanshahi and I. Kaastra. "A Comparison of Artificial Neural Network and Time Series Models for Forecasting Commodity Prices," Neurocomputing 10(1996):169-181.

Mahmoud, E. "Accuracy in Forecasting: a Survey," Journal of Forecasting 3(1984):139-159.

McClave, James T. and Dietrich II, Frank H., Statistics, Fourth Edition, San Francisco: Dellen Publishing Company, 1988.

Mendelsohn and Stein. "Fundamental Analysis Meets the Neural Network," Journal of Futures Markets 20(10):22-24, (1991)

Montgomery, Douglas C. Design and Analysis of Experiments. Third Edition. New York; John Wiley and Sons, 1991.

Ntungo, Chrispin. "Forecasting Performance Using Various Forecasting Steps Ahead Using Arima Models", P.H.D. Thesis, 1996.

Ntungo, Chrispin. "Commodity Futures Trading Performance Over Various Steps Ahead Forecasts Using Arima And Neural Network Models", P.H.D. Thesis, 1996.

Schwager, J. D. A Complete Guide to the Futures Markets. New York: John Wiley and Sons, 1984.

Tukey, J.W. "One Degree of Freedom for Non-Additivity." Biometrics, Volume 5, pp. 232-242, 1949.

APPENDIX A

The Two-Factor Factorial Design With One Observation Per Cell

The method outlined by Montgomery (1991) details the computations involved in performing an analysis of variance with two factors and only one observation per cell. Recall the linear statistical model is given by:

$$Y_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ij} \quad \{i=1,2,3,\ldots,a\} \; \{j=1,2,3,\ldots b\}$$

where Y_{ij} is the response variable, μ is the overall mean effect, τ_i is the effect of the ith level of the row factor A, β_j is the effect of the jth level of column factor B, $(\tau\beta)_{ij}$ is the interaction term and ε_{ij} is the random error component.

We assume for simplicity sake that the interaction term is zero which allows us to have a residual mean square which is an unbiased estimator of σ^2. The residual sum of square is partitioned into a single-degree-of-freedom component due to nonadditivity (interaction) and a component for error with (a-1)(b-1)-1 degree of freedom. Computationally we have

$$SS_N = \frac{[\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}y_{i.}y_{.j} - y_{..}(SS_A + SS_B + y_{..}^2/ab)]}{abSS_ASS_B}$$

with one degree of freedom, and $SS_{error} = SS_{residual} - SS_N$

with (a-1)(b-1)-1 degree of freedom.

The analysis of variance table is then formed and is outlined below as follows:

Table 4.1 Analysis of Variance for a Two-Factor Model, One Observation per Cell

Source of Variation	Sum of Squares	Degrees of Freedom	MeanSquare
Rows (A)	$\sum_{i=1}^{a} \frac{y_{i.}2}{b} - \frac{y_{..}2}{ab}$	a-1	MS_A
Columns (B)	$\sum_{j=1}^{b} \frac{y_{.j}2}{b} - \frac{y_{..}2}{ab}$	b-1	MS_B
Residual or AB	Subtraction	(a-1)(b-1)	$MS_{Residual}$
Total	$\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}2 - \frac{y_{..}2}{ab}$	ab-1	

The F-statistics are simply evaluated by dividing each of the mean square errors by the residual error. For example the F-statistic for the row factor A is MS_A divided by $MS_{Residual}$. If this value is greater than the tabled value of F with (a-1) degrees of freedom in the numerator and (ab-1) degrees of freedom in the numerator then we can conclude there is evidence to indicate that the Factor A is significant.

APPENDIX B

The general form to analyze a contingency table to test for independence is outlined in a text by (McClave and Dietrich, 1988). The null hypothesis is:

H_0: The two classifications are independent

H_1: The two classifications are dependent

The value of the test statistic χ^2 is computed using the formula:

$$\chi^2 = \sum \frac{\left[n_{ij} - \hat{E}(n_{ij})\right]^2}{\hat{E}(n_{ij})}$$

where

$$\hat{E}(n_{ij}) = \frac{r_i c_j}{n}$$

r_i is the total of the ith row and c_j is the total of the jth column.

The rejection region is: $\chi^2 > \chi^2_\alpha$, where χ^2_α has (r-1)(c-1) degrees of freedom.

APPENDIX C

Table 2.10 Analysis of Variance for the mean square error results of ARIMA Model Lag 1, Weekly Wheat Futures Closing Prices, Chicago Mercantile Exchange, 1976-1995

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F-Value
Forecast Interval	18254704	3	6084901.3	227.2
Sample Size	169576	2	84788.0	3.17
Nonadditivity	43136	1	43136	1.61
Error	133929	5	26785.8	
Total	18601345	11		

Table 2.11 Analysis of Variance for the mean square error results of ARIMA Model Lag 6, Weekly Wheat Future Closing Prices, Chicago Mercantile Exchange, 1976-1995

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F-Value
Forecast Interval	18909851	3	6303283.7	205.0
Sample Size	214033	2	107016.5	3.48
Nonadditivity	82624	1	82624	2.69
Error	153759	5	30751.8	
Total	19360267	11		

Table 2.12 Analysis of Variance for the mean square error results of ARIMA Model Lag 1, Weekly Cattle Futures Closing Prices, Chicago Board of Trade, 1976-1995

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F-Value
Forecast Interval	274.631	3	91.544	11.17
Sample Size	22.263	2	11.132	1.36
Nonadditivity	17.152	1	17.152	2.09
Error	40.969	5	8.194	
Total	355.015	11		

Table 2.13 Analysis of Variance for the mean square error results of ARIMA Model Lag 6, Weekly Cattle Futures Closing Prices, Chicago Board of Trade, 1976-1995

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F-Value
Forecast Interval	253.520	3	84.507	17.275
Sample Size	19.438	2	9.719	1.987
Nonadditivity	13.568	1	13.568	2.774
Error	24.458	5	4.892	
Total	310.984	11		

Table 2.14 Wheat Futures Results of Merton's Test of Turning Point Forecasting Power For Arima Models, and a Neural Network Model, 1976-1995

Model	α_0	α_1	R^2	RAF[1] (%)
Arima Lag 1	0.491	0.022	.0005	51.09
Arima Lag 6	0.521	-0.017	.0003	49.20
Neural Network	0.526	-0.024	.0006	48.82

[1] RAF stands for the ratio of accurate forecasts.

Table 2.15 Cattle Futures Results of Merton's Test of Turning Point Forecasting Power For ARIMA Models, and a Neural Network Model, 1976-1995

Model	α_0	α_1	R^2	RAF[1] (%)
Arima Lag 1	0.738	-0.077(*)	.0071	47.15
Arima Lag 6	0.563	-0.052	.0027	47.59
Neural Network	0.532	-0.007	.00005	49.78

[1] RAF stands for the ratio of accurate forecasts.

* Significant at the 5 percent level with the critical t- value of 1.645

Table 2.16 χ^2 **Test of Independence of Actual Value versus Forecast Value Results for ARIMA Models, and a Neural Network Model, Weekly Futures Prices, 1976-1995**

CATTLE 1975-1995

Model	Null hypothesis	Test statistic	Critical value 5%
Arima Lag 1	Forecast value and actual value are independent	4.845 *	3.841
Arima Lag 6	Forecast value and actual value are independent	1.856	3.841
Neural Network	Forecast value and actual value are independent	0.037	3.841
WHEAT 1975-1995			
Arima Lag 1	Forecast value and actual value are independent	0.327	3.841
Arima Lag 6	Forecast value and actual value are independent	0.187	3.841
Neural Network	Forecast value and actual value are independent	0.393	3.841

* Significant at the 5% level

Table 2.17 Test of the Difference of ratio of accurate forecasts for ARIMA Models versus a Neural Network Model, Weekly Futures Prices, 1976-1995.

CATTLE 1975-1995

Null hypothesis	Test statistic	Critical value 5%
$RAF_{c1}=RAF_{nn}$	0.974	1.645
$RAF_{c6}=RAF_{nn}$	0.811	1.645

WHEAT 1975-1995

$RAF_{w1}=RAF_{nn}$	0.838	1.645
$RAF_{w6}=RAF_{nn}$	0.140	1.645

All test statistics are insignificant at the 5% level
RAF_{c1} - The ratio of accurate forecasts value for the cattle lag 1 model
RAF_{c6} - The ratio of accurate forecasts value for the cattle lag 6 model
RAF_{w1}- The ratio of accurate forecasts value for the wheat lag 1 model
RAF_{w6}- The ratio of accurate forecasts value for the wheat lag 6 model
RAF_{nn} - The ratio of accurate forecasts for the neural network model