Neural Network Forecasting with the S&P 500 Index Across Decades

Mary Malliaris[1] and A.G. Malliaris[2]
[1]Information Systems & Operations Management, Loyola University, Chicago, IL, USA
[2]Economics, Loyola University, Chicago, IL, USA

Abstract - *The purpose of this paper is to track the effectiveness of a neural network as a forecasting tool across six decades, using only information derived from closing prices. From 1950 through 2010, a neural network for each decade was trained on ten years of S&P 500 data and used to forecast the S&P 500's direction each day of the following year. The set of inputs and structure of the networks remained constant across time. Only the data sets used for training and forecasting changed. The results show that, with one exception over 60 years, the neural networks remained robust from training to validation sets and were correct more than 50% of the time.*

Keywords: Data Mining, S&P 500 Index, Financial Forecasting, Neural Networks

1 Introduction

Many different models have been used in predicting the S&P 500 stock index and its behavior. Some models use technical indicators and others add fundamental indicators or economic growth indicators. When neural networks are used, the focus is often on a short period of time with a network optimized for that period. The aim of this paper is to develop a neural network using only data based on the closing value of the S&P 500 Index and then apply it to six decades of data, using the same structure and set of inputs across decades. The objective is to see whether this small network with no outside information can be a viable guide for a trading strategy.

From the early 1970s, literature on the behavior of stock prices has been divided between theories supporting market efficiency and active portfolio management. Proponents of market efficiency believe that information is incorporated quickly into the market and that prices fully reflect this information. As a result, prices cannot be predicted because they are changed by the constant arrival of new information. Traders, on the other hand, maintain a belief that forecasting is possible. However, only a few of them have managed to outperform the market over decades. Thus, while market efficiency remains the dominant theory, much effort is expended both by money managers and academics in an effort to predict well over time. One common support on the side of trading comes from the use of technical analysis.

A varied sample of studies over the years that have examined the usefulness of technical analysis and active management strategies include [1], [2], [3], [4], [5], [6], [7], and [8]. These studies span the spectrum of findings. Some are critical of simple technical rules and find the random walk does as well; others find that, once transactions costs are included, the predictive advantage of trading rules is moderated; and finally, some find evidence that some technical indicators have significant ability to aid in predictions.

In a recent paper, Schulmeister [9] looked at technical trading strategies on the S&P500 futures and their ability at predicting returns. This paper found that, in the 1960s and 1970s, the use of daily stock data was profitable. But the same indicators from 2000-2006 had lessening results for those strategies. One possible explanation given by the author is that the trend to higher frequency in trading on technical indicators gives insufficient time to produce a profitable strategy.

Other papers have focused more on fundamental aspects and macroeconomic data when developing forecasting models. Doran, Ronn, Goldberg [10] found that short term expected returns were highly volatile. Avramov and Chordia [11] used firm specific factors to predict returns. Prominent factors for predicting S&P returns are the Treasury yield and dividend yield. However, this predictability holds best for small-cap stocks, growth stocks, and momentum stocks, and not the broader market. Hajizadeh,et al [12] used Garch and neural network models to successfully forecast the S&P volatility. Niaki and Hoseinzade [13] looked at 27 potential financial and economical variables from March 1994 through June 2008 and were successful forecasting using this large set of internal and external variables. Fukushima [14] followed a number of hybrid models on monthly data and recommended complex hybrid models as the best method for forecasting. Tsiah et al [15] had earlier developed a hybrid neural network and rule-based system that predicted effectively over a six-year period. This paper develops a number of specialized signals similar to those of technical analysis. Kara et al [16] used ten technical indicators as inputs in both an artificial neural network model and a support vector machines model and found that the ANN outperformed the SVM.

In this paper, rather than using technical indicators, fundamental aspects, or macroeconomic data, we build a

network using variables derived only from the S&P 500 index daily prices. We then investigate the ability of this single network structure to forecast for over six decades. The next section describes the data and the network used. Section 3 details the results from this set of neural networks. We end with conclusions and recommendations for further research.
.

2 Data and Network Description

The data set began with the raw closing values of the S&P 500 from 1950 through 2010. These raw values were used to construct the other fields used as inputs. From the closing values, we calculated the percent the closing value changed, a four-day moving average of the closing values, and the percent change in these moving averages.

We then looked at the type of movement each day from the previous day, and logged it as having gone up or down. A string of two-day movement was formed by concatenating today's direction with yesterday's direction. For example, if the S&P 500 moved up yesterday and down today, the string UD was entered. In a similar fashion, strings of three, four and five days up and down movements were recorded. In Figure 1, we show, as an example, the percent of time over each decade that the possible four-day strings, DD, DU, UD, and UU have occurred. One interesting pattern we see is that, in every decade except the last one, the most often occurring string was UU. We also see that the string DD was on the rise from the 50s through the 70s, then decreased. Lastly, we see the increase over every decade of movement shifts. That is the percent of time that UD and DU occur increases from the 50s through the 00s. Charts for three, four, and five day patterns also indicate similarities in dominant patterns over the decades.

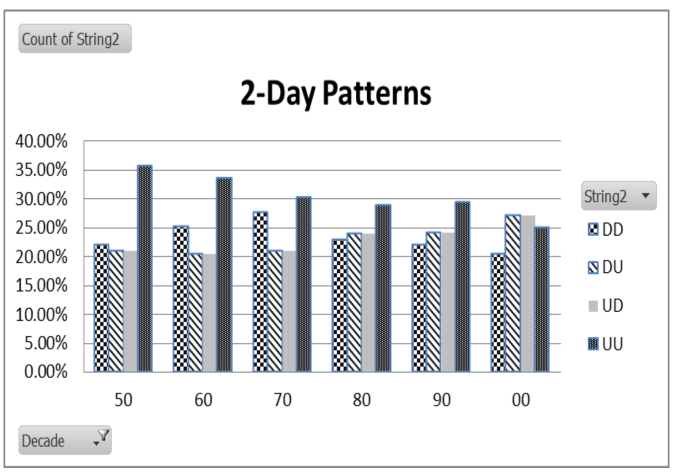

Fig 1 The four two-day strings of Up and Down across decades

Another ways of giving information to the networks is by condensing these directional movements to a count of the number of Up movements in strings of a given length. So we counted the number ups in strings of length 1 to 5 and used these as additional inputs. That is, our focus shifted from the exact pattern to a count of positive moves within a specific number of days. Figure 2 shows the result of converting the three days strings into this type of count. Within three days, it is possible to have 0, 1, 2, or 3 up movements. Looking at these counts across the decades, we see that the percent of times that three days in a row were all up has steadily decreased from the 50s through the 00s. In addition, there are more occurrences with exactly 2 ups than with exactly 1 up with three days. Last, the percent of times that three days were all down has been decreasing since the 70s.

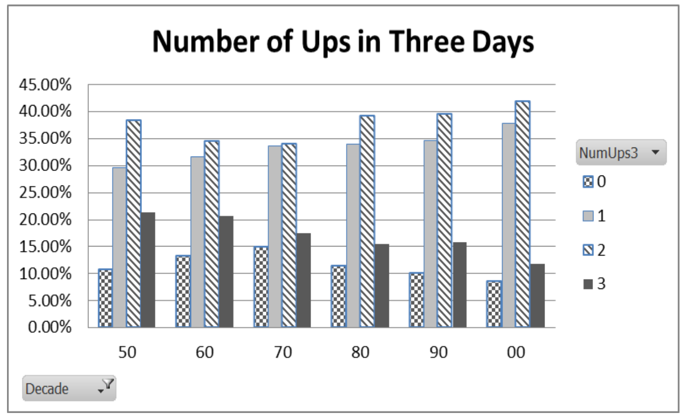

Fig 2. Count of Up Movements in Three Days, shown as percent within each decade.

The entire set of 14 inputs used in each of the networks is listed in Table 1. This table has a column with the label used for the input, an explanation of that input, and a sample value. The Target field, DirTp1, not shown in the table, was a prediction of the direction the S&P will move tomorrow, Up or Down.

After the columns were constructed for the entire data set, subsets were formed for the training and validation sets. A training set was fashioned for each of the decades where the entire ten years of data was available. We had a total of six training sets for the decades from 1950 through 2009. Validation sets were comprised of the entire year immediately following the associated training set. Specific dates for each of the training and validation sets are shown in Table 2.

All networks were developed and run in IBM's SPSS Modeler 14 software package. This package automatically selects an optimal network structure and settings. However, networks with alternate structures were also tested. Using the same inputs, networks with the recommended hidden layer of 9 nodes were tested against networks with hidden layers of 14 nodes and 20 nodes. The networks with hidden nodes of equal size and fan-out size did not improve the performance, so we used the Modeler suggested form with one hidden layer of nine nodes. Thus, for each multilayer perceptron, there were 14 inputs, one hidden layer with 9 nodes, and one output.

Table 1. Inputs for each network

Input	Explanation	Example
Close	Today's Closing Value	1132.99
PercChgClose	Percent Change in the Closing Value	1.60
CloseDir	Today's Closing Direction	U
MA4day	4 Day Moving Average of Closing	1137.08
PercChg4MA	Percent Change in the 4-day Mov. Avg.	-0.0593
NumUps1	Was today's close an Up move	1
NumUps2	Number of Up Closings in last 2 days	1
NumUps3	Number of Up Closings in last 3 days	2
NumUps4	Number of Up Closings in last 4 days	2
NumUps5	Number of Up Closings in last 5 days	3
String2	2-day Up and Down pattern	DU
String3	3-day Up and Down pattern	UDU
String4	4-day Up and Down pattern	DUDU
String5	5-day Up and Down pattern	UDUDU

Table 2. Data sets for each network.

Decade	Training/Testing Set	Validation Set
50s	Jan 1, 1950 -- Dec 31, 1959	Jan 1, 1960 – Dec 31, 1960
60s	Jan 1, 1960 -- Dec 31, 1969	Jan 1, 1970 – Dec 31, 1970
70s	Jan 1, 1970 -- Dec 31, 1979	Jan 1, 1980 – Dec 31, 1980
80s	Jan 1, 1980 -- Dec 31, 1989	Jan 1, 1990 – Dec 31, 1990
90s	Jan 1, 1990 -- Dec 31, 1999	Jan 1, 2000 – Dec 31, 2000
00s	Jan 1, 2000 -- Dec 31, 2009	Jan 1, 2010 – Dec 31, 2010

With this 14-9-1 structure, the value used for the random seed was 229176228 and 30% of the training set was used to prevent over-fitting. The training algorithm used by Modeler stops after 15 minutes, or when the error in the over-fit prevention set does not decrease after each cycle, if the relative change in the training error is small, or if the ratio of the current training error is small compared to the initial error. The structure of a typical network used for each of the decades is shown in Figure 3.

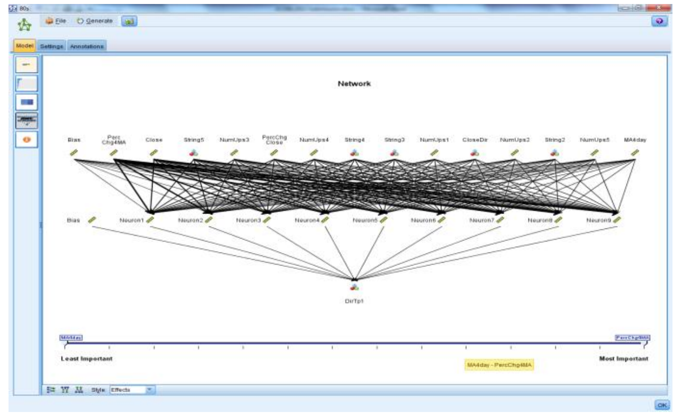

Fig 3. Structure of each of the neural networks.

3 Results

After training a network for each decade, Modeler allows us to look at the results in several ways. First, Modeler displays the 10 variables that have the greatest significance in determining the final value of the target. This is called the predictor importance, and indicates the relative importance of each predictor in estimating the model. All values assigned to these variables are relative to the variable's impact and their numeric values sum to 1.0. Predictor importance does not relate to model accuracy. It is the importance of each predictor in making a prediction, not whether the prediction is accurate. Predictor importance is calculated from the test partition and looks at the impact each variable has on changes in the target field. The relative importance of the ten highest variables for each network is shown in Table 3. Here we see that the percent change in the closing price and in the 4-day moving averages are highly ranked in most decades. We also see that the string of 5 days has a lot of impact every decade. In addition, the number of up movements within sets of four days occurs in many of the network lists.

In the outputs from Modeler, a matrix showing the count of correct and incorrect predictions is generated. We can also feed other sets through the trained network to generate counts of prediction accuracy on new data sets. Table 4 shows the overall percent of times that the network was correct on the training and validation sets. We see, in the decades of the 50s, 60s, and 70s, both the training and validation sets are correct close to sixty percent of the time. The 80s, 90s, and 00s show a drop in the overall ability to forecast with this methodology, but with the exception of the final validation set, all results are still better than 50%.

Table 5 breaks these forecasts down further into each direction. The rows show actual Down and Up values while the columns have the predicted Down and Up movements. The percent of predictions correctly matching the actual values are in the diagonal and shown in bold. The off-

Table 3. Relative Importance of Top Ten Variables in Each Network.

50s	60s	70s	80s	90s	00s
PercChgClose	PercChg4MA	PercChgClose	PercChg4MA	PercChg4MA	PercChg4MA
PercChg4MA	NumUps4	PercChg4MA	Close	NumUps5	NumUps4
String5	PercChgClose	Close	String5	PercChgClose	MA4day
String4	String4	MA4day	NumUps3	MA4day	String5
NumUps1	String5	NumUps2	PercChgClose	String3	Close
NumUps4	NumUps5	String5	NumUps4	String5	PercChgClose
String3	String2	String3	String4	String4	String4
MA4day	String3	CloseDir	String3	Close	String3
NumUps5	MA4day	String4	NumUps1	String2	NumUps1
NumUps3	Close	NumUps4	CloseDir	CloseDir	NumUps3

Table 4. Percent of Correct Forecasts in Training and Validation Sets

Decade	Tr Percent Correct	Val Percent Correct
50s	59.00%	59.92%
60s	59.90%	62.99%
70s	59.94%	58.10%
80s	55.18%	54.94%
90s	56.25%	52.78%
00s	52.96%	47.22%

Table 5. Comparison of Training Set and Validation Set Performance, Values as % of Column

		Training Set			Validation Set	
		Predictions			Predictions	
Actual Direction		**Down**	**Up**		**Down**	**Up**
50s	**Down**	**53.90**	39.05		**61.68**	41.38
	Up	46.10	**60.95**		38.32	**58.62**
60s	**Down**	**57.94**	39.00		**67.02**	39.38
	Up	42.06	**61.00**		32.98	**60.62**
70s	**Down**	**58.88**	39.06		**53.25**	39.77
	Up	41.12	**60.94**		46.75	**60.23**
80s	**Down**	**53.21**	43.84		**63.16**	45.73
	Up	46.79	**56.16**		36.84	**54.27**
90s	**Down**	**53.43**	42.02		**56.70**	49.68
	Up	46.57	**57.98**		43.30	**50.32**
00s	**Down**	**50.67**	44.99		**40.74**	45.30
	Up	49.33	**55.01**		59.26	**54.70**

diagonal numbers indicate the percent of incorrect predictions. For example, the training set of the 50s correctly predicted Down 53.9% of the time, and correctly predicted Up 60.95% of the time. The validation set used on this network correctly predicted Down 61.68% of the time, and Up predictions were correct 58.62% of the time.

For the training set data, we see that the percent of correct Up forecasts is greater than the percent of correct Down forecasts in every decade, even though there is a slight decrease over time in these values. In contrast, the percent of correct validation set forecasts are greater for the Down forecasts in four out of six decades. In particular, the last validation set, which had less than 50% accuracy overall on the validation set, turns out to do much better on the Up forecasts. It is only in trying to predict the Down days that the network falls below 50% correctness.

4 Conclusions

In this paper, we built a series of neural networks using information constructed only from the closing values of the S&P 500 Index. These networks covered over sixty years and included a training set for each decade followed by a one year validation set from the following decade. All networks used the same 14-9-1 topology, the same random seed, and a testing set with 30% of the data to prevent overtraining. In addition, each trained network was applied to a validation set of the entire following year. There were fourteen input variables based on the closing values and direction of movement in comparison to the previous day. From among the fields calculated by using the numeric closing values, those with greatest impact were the percent change in the closing price relative to yesterday and the percent change in the four-day moving average of closing prices. From among the up and down string patterns, the five-day pattern had the most consistent impact. Last, from the fields that counted the number of up days in strings of a given length, the four day count appeared higher up on the list in most of the networks. In every decade, the networks did better than 50% correct predictions on both training and validation sets, except in the last validation set. In this last set, the percent of correct forecasts in the up direction was almost 55%, while the down forecasts were correct only 41% of the time.

Other than retraining the network on each decade, no other changes were made to the neural network, and all information given to the network came from variables constructed using the daily closing price. It is interesting that this identical structure, using the same inputs, was useful for over six decades. Future research might investigate a smaller training time, say a rolling window of one or two years. This might enable us to see the importance of specific variables gradually shifting over time.

5 References

[1] Fama E, Blume M. « Filter Rules and Stock-Market Trading » ; J of Bus 39:226-241, 1966.

[2] Jensen M, Benington G. « Random Walks and Technical Theories: Some Additional Evidence » ; J of Fin 25:469-482, 1970.

[3] Brown D, Jennings R. « On Technical Analysis » ; The Rev of Fin Studies, 2:527-551, 1989.

[4] Brock W, Lakonishok J, LeBaron B. « Simple Technical Trading Rules and the Stochastic Properties of Stock Returns » ; J of Fin, 47:1731-1764, 1992.

[5] Blume L, Easley D, O'Hara M. « Market Statistics and Technical Analysis: The Role of Volume » ; J of Fin, 49:153-181, 1994.

[6] Gencay R « The Predictability of Security Returns with Simple Technical Trading Rules » ; J of Empirical Fin, 5:347-359, 1998.

[7] Allen F, Karjalainen R « Using Genetic Algorithms to Find Technical Trading Rules » ; J Fin Econ, 51:245-272, 1999.

[8] Lo A, Mamaysky H, Wang J. Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation. J of Fin 55:1705-1765, 2000.

[9] Schulmeister S. « Profitability of technical stock trading: Has it moved from daily to intraday data?, Review of Financial Economics, 18:4, pp : 190-201, October 2009.

[10] Doran J, Ronn E, Goldberg R. « A Simple Model for Time-Varying Expected Returns on the S&P 500 Index » ; Journal of Investment Management, 7 :47-72, 2009.

[11] Avramov D, Chordia T. « Predicting Stock Returns » ; Available at SSRN: http://ssrn.com/abstract=352980 or http://dx.doi.org/10.2139/ssrn.352980, March 23, 2005.

[12] Hajizadeh E, Seifi A, Fazel Zarandi M, Turksen I. « A hybrid modeling approach for forecasting the volatility of S&P 500 index return » ; Expert Systems with Applications, Vol 39:1, pp: 431-436, 2012.

[13] Niaki S, Hoseinzade, S. « Forecasting S&P 500 index using artificial neural networks and design of experiments » ; Journal of Industrial Engineering International, Vol 9:1, 2013.

[14] Fukushima A. « Hybrid forecasting models for S&P 500 index returns » ; The Journal of Risk Finance, Vol. 12 (4), pp.315 – 328, 2011.

[15] Tsaih R, Hsu Y, Lai C. « Forecasting S&P 500 stock index futures with a hybrid AI system » ; Decision Support Systems, 23 :2, pp: 161-174, 1998.

[16] Kara, Y., Acar Boyacioglu, M., & Baykan, Ö. K. « Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange » ; Expert Systems with Applications, Vol 38(5), 5311-5319, 2011.