

Forecasting stock index returns using a volatility based recurrent neural network *

Stelios D. Bekiros & Dimitris A. Georgoutsos

**Department of Accounting and Finance
Athens University of Economics and Business
76 Patission str.,
104 34 Athens, GREECE**

October 2004

Abstract

This paper investigates the profitability of a trading strategy, based on recurrent neural networks, that attempts to predict the direction of the market in the case of the NASDAQ general index. The sample extends over the period 2/8/1971 – 4/7/1998, while the sub-period 4/8/1998 – 2/5/ 2002 has been reserved for out-of-sample testing purposes. We demonstrate that the incorporation in the trading rule of estimates of the conditional volatility changes strongly enhances its predictability during “bear” market periods. This improvement is being measured with respect to a nested model without the volatility variable as well as to a buy & hold strategy. We suggest that our findings can be justified by invoking either the “volatility feedback” theory or the existence of portfolio insurance schemes in the equity markets.

(*) preliminary version; not to be quoted

Keywords: Technical trading rules; Neural network models; Security markets

JEL classification: G10; G14; C53

1. INTRODUCTION

Over the past years an important line of research has focused on testing the predictability of stock returns. Overall, the findings of this literature are that stock returns are predictable from past returns and other economic and financial variables. For instance, Fama and French (1988) and Poterba and Summers (1988) assume that stock returns are generated by the sum of a random walk and a stationary mean-reverting process and report that long (three-to five-year) holding period returns are significantly negatively serially correlated. Lo and MacKinlay (1988) use a volatility based specification test and reject the random walk hypothesis in favor of positive serial correlation for weekly and monthly holding-period returns of value weighted and equally weighted portfolios. Then they show that nonsynchronous trading cannot fully account for the observed autocorrelation while they further report slightly negative correlation for individual securities that is attributed to their considerable idiosyncratic noise. Lo and MacKinlay (1990) attempt to reconcile the positive serial dependence in market indexes returns with the negative autocorrelation in individual returns. They show that positive cross-autocorrelations across securities can account for more than 50 percent of the profits from contrarian investment strategies. Pesaran and Timmermann (1995) examine whether the predictability of the Standard and Poor's 500 index returns could have been historically exploited by investors to earn profits in excess of a buy-and-hold strategy. They also find that the predictive power of various economic factors is increased during volatile periods.

The focus of the papers cited above, and in many others in the related literature, is the linear predictability of asset returns that relies on weighted

combinations of return autocorrelations. However, many aspects of economic behavior may not be linear. Recent research into the time series properties of stock market indices by Pesaran and Timmermann (1994) and Abhyankar, Copeland and Wong (1997), to name a few, have indicated the presence of non-linear dynamics. An interesting issue therefore that is worth addressing is whether the predictability of stock returns can be improved by using non-linear models. One of the approaches that have been found to improve the ability of forecasting asset returns is the Artificial Neural Networks (ANNs). An ANN can be viewed as a general nonlinear time series model that under general regularity conditions can approximate any number of a class of functions to any desired degree of accuracy. Gençay (1998) incorporates the buy and sell signals from a simple technical trading strategy, which is based on a moving average rule, into an ANN specification. The results show that the moving average rules in an ANN provide a forecast improvement, as measured by the mean square prediction errors (MSPEs), when they are compared with the predictions of the Dow Jones index daily returns from a linear regression or a GARCH-M(1,1) process. Gençay and Stengos (1998) extend the previous work by incorporating a volume average rule in their trading strategy. The MSPEs of the ANN conditional mean specification are smaller than those obtained from the benchmark, linear and GARCH-M (1, 1) models. Gençay (1998) extends the evaluation criteria of the competing strategies to include measures of profitability. The results indicate that nonparametric models with technical strategies provide significant profits when tested against a simple buy-and-hold strategy. Fernández-Rodríguez *et. al.* (2000) conduct a similar exercise for the Madrid stock market general index and show that a simple trading rule based on ANNs is always superior to a buy-and-hold strategy during “bear” market conditions.

In the present we explore the predictive ability of trading rules that incorporate, among others, estimates of the conditional volatility changes over the next trading period. The empirical investigation of the relation between stock return volatility and stock returns has a long tradition in finance. According to the “time-varying risk premium theory” the return shocks are caused by changes in conditional volatility. When news arrives in the market the current volatility increases and this causes upward revisions of the conditional volatility since there is a well-documented fact that volatility is persistent. This increased conditional volatility has to be compensated by a higher expected return, leading to an immediate decline in the current value of the market. So in the case of bad news the volatility feedback effect reinforces the initial drop in stock market prices. However when good news arrives in the market and volatility increases, prices decline to induce higher expected returns offsetting thus the initial price movement.¹ An alternative rationalization for the presence of conditional volatility revisions in the trading rule may be offered by invoking trigger strategies in the equity markets (see, e.g. Krugman, 1987). When participants in portfolio insurance schemes react whenever the maximum expected loss, as measured by the Value-at-Risk (VaR), reaches a predetermined level then we will observe share price dynamics that are being driven, partly, by revisions in the measured conditional volatility.² If we assume a continuity of portfolios that deviate to a varying degree from their pre-determined level of VaRs then each time the conditional volatility rises, a number of those portfolios will hit their risk limits and

¹ An asymmetric nature of the volatility response to return shocks emerges from the above discussion. Bad news generates an increase in conditional volatility while the net impact of good news is not clear. An alternative explanation to the asymmetric reaction of the conditional volatility may be offered through the “leverage effects” [see, e.g. Christie (1982)]. A negative (positive) return increases (reduces) financial leverage, which makes the stock riskier (less risky) and increases (reduces) volatility. The causality however here is different: the return shocks lead to changes in conditional volatility, whereas the time-varying premium theory contends the opposite [see Bekaert and Wu, 2000].

² The VaR depends entirely on a multiple of the estimated conditional volatility under the assumption of normally distributed returns.

this will generate a re-allocation of assets towards safer ones. Each time portfolio insurers leave the market the stock prices must fall in order for the other investors to be given an incentive to hold a larger quantity of stock. If we further assume a rational expectations world then investors take into account the effects of portfolio insurance schemes and no step drop in stock prices is being observed.³

The structure of the paper is as follows. In section two we discuss the construction of the trading rule and the way this is incorporated into an ANN as well as the estimation techniques that have been applied. In section three we present the data, we discuss the statistical and financial criteria we adopt to evaluate the forecasting ability of the various models and finally the empirical results are shown. Section four provides some concluding remarks.

1. METHODOLOGY

Neural networks use a non-parametric method of forecasting which means that the underlying non-linear function is not prescribed, ex-ante, explicitly. Thus, the model is not limited to a restrictive list of non-linear functions.⁴ In financial applications the most popular class of ANN models has been the single-layer feedforward networks (FNR). In a FNR, information, suitably weighted, is passed from the point of entry (the input layer) to a further layer of hidden neurons. This hidden information is also assigned a weight and finally reaches the output layer which represents the forecast.

³ Under static expectations however the portfolio insurance schemes can cause a stock market crash as the market moves to its new regime (see Krugman, 1987).

⁴ Cybenko (1989) and Hornik *et.al.* (1989) have demonstrated that ANN models can approximate, under certain regularity conditions, any continuous function. This unveils the main weakness of the ANNs since they may end up fitting the noise in the data rather than the underlying statistical process. Cheng and Titterton (1994) have shown that ANNs are equivalent to non-linear non-parametric models while they claim that most forecasting models (ARMA, autoregressive with thresholds, non-parametric with kernel regression, etc.) can be written in the form of a network of neurons.

Let $p_t, t = 1, 2, \dots, T$ be the daily stock index price. The daily returns are then calculated by $r_t = \log(p_t) - \log(p_{t-1})$. The output, y , of a single layer FNR is then given by:

$$y_t = S \left[a_0 + \sum_{i=1}^q a_i \cdot g_{t,i} \right] \quad (1)$$

where:

$$g_{t,i} = G \left(b_{i0} + \sum_{j=1}^n b_{ij} r_{t-j} + \sum_{k=0}^p c_{ik} \Delta h_{t-k}^{1/2} \right), \quad i = 1, \dots, q. \quad (2)$$

The inputs in the suggested FNR correspond to the returns in the previous n days, following Gençay (1999) and Fernandez-Rodriguez *et.al.* (2000), and the revisions of the estimated conditional volatility, $\Delta h^{1/2}$, over the past p days. As concerns the transfer functions G and S we use the tansig and the purelin function respectively. The tan-sigmoid function normalizes the values of each neuron to be in the interval $(-1, +1)$ while the linear output layer lets the network produce values outside the range -1 to $+1$. The problem we are faced with is that the FNR has the correct weights such that y has the correct value corresponding to the inputs. This is being accomplished by the *error backpropagation* method under which the neural network runs through all the input data over an initial “training” period and produces a list of outputs. Then the weights are reevaluated, by using a recursive “gradient” descent method, so that the mean-squared error between the observed output and the predicted one is minimized.⁵ Once the neural network has been trained, it is applied over a different data set

⁵ As White (1992) has shown the existence of such weights is guaranteed since any non-linear function can be approximated as above, with a single layer, to an arbitrary degree of accuracy with a suitable number of neurons, g .

covering the so-called “validation” period. The purpose here is to evaluate the generalization ability of a supposedly trained network in order to avoid overfitting.

In a dynamic context it is natural to include lagged dependent variables as explanatory variables in the FNR in order to capture dynamics. This problem is being addressed in the relevant literature by constructing recurrent networks, i.e. networks with feedbacks from the hidden neurons, g , to the input layer with delay. The recurrent neural networks (RNR) memorize thus information since its output depends on both current and prior inputs. In this paper we apply the Elman (1990) RNR with a single hidden layer and feedback connection from the output of the hidden layer to its input. In a RNR model equation (2) can be re-written as:

$$g_{t,i} = G \left(b_{i0} + \sum_{j=1}^n b_{ij} r_{t-j} + \sum_{k=0}^p c_{ik} \Delta h_{t-k}^{1/2} + \sum_{l=1}^m \delta_{il} g_{l,t-1} \right), \quad i = 1, \dots, q \quad (3)$$

and it is easy to show, with back-substitution, that the output y_t depends on the entire history of the inputs r and $\Delta h^{1/2}$.

The trading rule over the testing period works as follows. At the end of each trading day the RNR is being re-estimated over a rolling sample that is equal to the training period set. The output unit, eq. (1), receives the weighted sum of the signals, in the $(-1, 1)$ interval, from eq. (3) and produces a signal through the output transfer function (S). If the value of the signal is greater than zero it is interpreted as a “buy” signal for the next trading day while a value less than zero as a “sell” signal. Then, the total return of the strategy, when transaction costs are not considered, is estimated as:

$$\hat{R}_t = \sum_{t=0}^N \hat{y}_t r_t, \quad (4)$$

where \hat{y}_t is the recommended position which takes the value of (-1) for a short position and (+1) for a long position (see e.g. Gençay, 1998a and Fernandez-Rodriguez *et.al.*, 2000).

3. EMPIRICAL RESULTS

We have estimated the RNR model of equations (1) and (3) on daily returns of the NASDAQ index that span the period 2/8/1971 to 2/5/2002 (figure 1). The testing, out-of-sample, period has been split into two subperiods; a “bull” market period from 4/2/1998 to 3/12/2000 and a “bear” market period from 3/1/2000 to 2/5/2002. The training and validation period account for the rest of the sample with the validation period covering almost 30% of the entire data set.

In order to rationalize the use of neural network models we have tested for the presence of non-linear dependence in the series. To that end, we have made use of the well known BDS test statistic which under the null of i.i.d. is given by (see Brock *et.al.*, 1991):

$$W_{m,T(\varepsilon)} = T^{1/2} [C_{m,T(\varepsilon)} - C_{1,T(\varepsilon)}] / \sigma_{m,T(\varepsilon)} . \quad (5)$$

$C_{m,T(\varepsilon)}$ is the correlation integral from m dimensional vectors that are within a distance ε from each other, when the total sample is T , and $\sigma_{m,T(\varepsilon)}$ is the standard deviation of $C_{m,T(\varepsilon)}$. Under the null hypothesis, $W_{m,T(\varepsilon)}$, has a limiting standard normal distribution. The BDS test has been applied on: (a) the original data, (b) the residuals from an autoregressive filter, in order to ensure that the null is not rejected due to linear dependence, and (c) the natural logarithm of the squared standardized residuals from a GARCH-M (1,1) model, in order to ensure that rejection of the null is not due to conditional heteroscedasticity (see De Lima, 1996). The results are shown in Table 1.

In all three cases we were unable to accept the null of i.i.d. at the 1% marginal significance level and the evidence seems to suggest that a genuine non-linear dependence is present in the data.

The results relating to the predictability as well as the profitability of the Elman network we estimated appear in table 2. They correspond to a specification where two lags of the returns and no lag of the conditional volatility changes appear in equation (3), ($n=2$ and $p=0$).⁶ In addition there is one hidden layer with ten neurons (g) and one output layer with a single neuron (y). Conditional volatility estimates, $h_t^{1/2}$, have been obtained from: a rolling 20-day standard deviation of returns; an exponentially weighted moving average with decay factor equal to 0.94;⁷ a GARCH (1,1) model; and a Glosten, Jagannathan and Runkle (1993) (GJR) GARCH (1,1) model that allows for an asymmetric response of volatility to positive or negative shocks.

The adequacy of the chosen specification, without the presence of the volatility changes, is considered as satisfactory. As the first column of table 2 shows the total return of the trading strategy is 29.2% for the entire testing period when a buy-and-hold (B&H) policy would have earned only 4.5%. Moreover, the proportion of the correctly predicted signs is above 50% and this is reflected in a significant value, at the 5% level, of the Henriksson- Merton (HM) test statistic. Finally, the Sharpe ratio (SR) and the Ideal profit (IP) index are both positive, although rather small in value. As concerns the two testing sub-periods, the chosen strategy behaves

⁶ The procedure for the selection of the lags involved the estimation of AR models and the calculation of the Ljung-Box statistics for the first 16 lags of the series. Significant autocorrelations of up to the second lag of the return series were identified. As concerns the conditional volatility variable, sensitivity analyses for different number of lags were conducted on the RNR but the results were not found to be qualitatively different from those presented in table 2. Similar exercises were conducted for a different number of lagged returns but again the results we obtained are not better than those shown in table 2. The results of the sensitivity analyses are available upon request.

⁷ The exponentially moommage corresponds to the approach adopted by RiskMetrics and for that reason it is denoted here as RM(0.94).

better during the bull period, according to the Pesaran-Timmerman (PT) and HM tests as well as the SR and IP index. However, the overall return compared to the B&H policy is superior during the “bear” market sub-period. This image accords with previous results derived by Fernández *et. al.* (1999, 2000) from a similar model applied on the Nikkei and the Madrid stock market General Indices.

Next, we evaluate the trading strategy with the conditional volatility variable included. The first evidence that emerges from this change is that the returns improve significantly, over the “bear” market period, independently of the model we used to produce the conditional volatility estimates. Over, the “bull” market period the strategy does not seem capable of succeeding the profits with the “no volatility” strategy while it is always below the B&H policy. Similarly, the other performance indices, PT, HM, SR and PI, show a substantial improvement over the “no volatility” case for the “bear” market period. Those same indices for the “bull” market period produce values that are similar or worse to those under the “no volatility” specification. The comparison between the four different specifications for the volatility estimation show that simple models of historical volatility measurement, like the equally weighted and the exponentially weighted moving averages, produce substantially better forecasts than the more complicated econometric models that are often used to model conditional volatility.⁸

4. CONCLUDING REMARKS

In the present paper we expand the literature that evaluates the forecasting ability of trading rules based on neural networks over simple alternative strategies like Buy & Hold. A B&H policy can not be consistently outperformed from any trading

⁸ This has been surprising since it is well documented that forecasts of volatility, for the NASDAQ index, from MA rules closely approximate those from GARCH (1,1) models (see e.g. Schwert, 2002).

rule, no matter how elaborate this is, in a random walk market. We first replicate previous evidence coming from other stock market indices, according to which the forecasting ability of simple rules outperform the B&H profits over “bear” market conditions although the evidence from various profitability indices is positive for the “bull” period as well. Then we included in the trading rule revisions of the conditional volatility of the NASDAQ index that have been produced from alternative estimating techniques. This change generated a substantial improvement of the profits, the market timing ability and the profitability per unit of risk over the “bear” market period. Those results seem to indicate that the neural network has been “trained” to relate correctly changes in conditional volatility with the “sign” of the market one day ahead. This may be attributed to two factors. The first associates increases in volatility to higher expected returns. In the case when increases in volatility have been generated from “bad” news we will definitely experience lower prices the next trading day. However, when increases in volatility are generated from “good” news it is not clear what the net effect on prices will be. This explanation seems to accord with the enhanced predictability of the model that incorporates the volatility revisions over the “bear” market period. Unfortunately however, when we adopt a model of asymmetric volatility the evidence in favor of the explanation offered above deteriorates. The second factor associates increases in volatility with trigger strategies followed by many portfolio managers. Each time volatility rises the risk limit is being hit for some portfolios and then liquidation follows. This puts a pressure on the market that is more severe during “bear” market conditions.

References

- Abhyankar, A., Copeland, L. S. and Wong, W., 1997, "Uncovering Nonlinear Structure in Real-time Stock Market Indexes: The S&P 500, the DAX, the Nikkei 225, and the FTSE-100, Journal of Business & Economic Statistics, 15, 1-14.
- Bekaert, G. and G. Wu, 2000, "Asymmetric Volatility and Risk in Equity Markets", The Review of Financial Studies, 13, no 1, 1-42.
- Brock, W., D. Hsieh and B. LeBaron, 1991, "Nonlinear Dynamics, Chaos and Instability", The MIT Press.
- Cheng, B., and D.M. Titterton, 1994, "Neural networks: a review from a statistical perspective", Statistical Science, 9, 2-54.
- Cybenko, G., 1989, "Approximations by superpositions of a digmoidal function", Mathematics of Control, Signals and Systems, 2 , 203-230.
- De Lima, P.J.F., 1996, "Nuisance parameter free properties of correlation integral based statistics", Econometric Reviews, 15, 237-259.
- Fama, E., and K. French, 1988, "Permanent and Temporary Components of Stock Prices", Journal of Political Economy, 96, 246-273.
- Fernández-Rodríguez, F., S. Sosvilla-Rivero, and M. D. Garca-Artiles, 1999, "Dancing with Bulls and Bears: Nearest-neighbour forecasts for the Nikkei index", Japan and the World Economy, 11, 395-413.
- Fernández-Rodríguez, F., C. Gonzalez-Martel and S. Sosvilla-Rivero, 2000, "On the Profitability of Technical Trading Rules based on Artificial Neural Networks: Evidence from the Madrid Stock Market", Economic Letters, 69, 89-94.
- Gençay, R., 1998a, "The Predictability of Security Returns with Simple Technical Trading Rules", Journal of Empirical Finance, 5, 347-359.

_____, 1998b, "Optimization of Technical Strategies and the Profitability in Security Markets", Economic Letters, 59, 249-254.

_____ and T. Stengos, 1998c, "Moving Average Rules, Volume and the Predictability of Security Returns with Feedforward Networks", Journal of Forecasting, 17, 401-414.

Glosten, L., R. Jaganathan and D. Runkle, 1993, "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks", The Journal of Finance, 48, 1779-1801.

Hagan, M.D. and M. Menhag, 1994, "Training feedforward networks with the Marquardt algorithm", IEEE Transactions on Neural Networks, vol. 5, no 6, 989-993.

Henriksson, R.D. and R.C. Merton, 1981, "On the Market timing and Investment Performance II: Statistical procedures for evaluating forecasting skills", Journal of Business, 54, 513-533.

Hornik, K., M. Stinchcombe and H. White, (1989), "Multi-layer feedforward networks are universal approximations", Neural Networks, 2, 359-366.

Krugman, P., 1987, "Trigger Strategies and Price Dynamics in Equity and Foreign Exchange Markets", NBER Working Paper no 2459.

Lo, A. W., and A.C. MacKinlay, 1988, "Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test", Review of Financial Studies, 1, 41-66.

_____, 1990, "When Are Contrarian Profits Due to Stock Market Overreaction?", Review of Financial Studies, 3, 431-468.

Pesaran, M. H., and A. Timmermann, 1994, "Forecasting Stock Returns. An Examination of Stock Market Trading in the Presence of Transaction Costs", Journal of Forecasting, 10, 335-367.

_____, 1995, "Predictability of Stock Returns: Robustness and Economic Significance", The Journal of Finance, 50, 1201-1228.

Poterba, J., and L. Summers, 1988, "Mean Reversion in Stock Returns: Evidence and Implications", Journal of Financial Economics, 22, 27-60.

Schwert, W., 2002, "Stock Volatility in the new millennium: how wacky is Nasdaq?", Journal of Monetary Economics, 49, 3-26.

White, H., (1992), Artificial Neural Networks: Approximation and Learning Theory, Cambridge, MA, Blackwell Publishers.

Table 1: BDS test

Series	$m=2$		$m=3$		$m=4$	
	$\varepsilon = 1$	$\varepsilon = 1.5$	$\varepsilon = 1$	$\varepsilon = 1.5$	$\varepsilon = 1$	$\varepsilon = 1.5$
OD	5.18 [*]	4.93 [*]	8.98 [*]	8.24 [*]	10.74 [*]	9.54 [*]
RAF	4.74 [*]	4.53 [*]	8.30 [*]	7.80 [*]	9.87 [*]	9.05 [*]
NLSNR	3.15 [*]	3.64 [*]	3.97 [*]	4.22 [*]	4.20 [*]	4.19 [*]

Notes: OD = original data (daily returns of the NASDAQ index),

RAF = residuals from an autoregressive filter,

NLSNR = natural logarithm on standardized normalized residuals.

m = the value of the dimension, ε = the number of standard deviations of the data.

Brock *et. al.* (1991) suggest that the standardized normal distribution is a good approximation of the finite sample distribution for a sample of 500 or more observations, values of the dimension m below 5 and values of the distance ε between 0.5 and 2 standard deviations of the data.

(^{*}) indicates significance at the 1% significance level (the critical value is 2.58).

Table 2: Out-of-sample tests. Testing period: 4/2/1998 – 2/5/2002. “Bull” market period: 4/2/1998 to 3/12/2000. “Bear” market period: 3/1/2000 to 2/5/2002.

	RNR (no volatility)			RNR - MA(20)		RNR - RM(0.94)	
<i>Sub-period</i>	Testing period	Bull	Bear	Bull	Bear	Bull	Bear
Total Return	0.292	0.569	-0.277	0.518	1.087	0.400	1.207
B&H Return	0.045	1.027	-0.982	1.027	-0.982	1.027	-0.982
Sign Rate	0.525	0.543	0.507	0.535	0.533	0.517	0.547
PT test	1.480***	1.665**	0.382	1.477***	1.513***	0.645	2.195**
Merton test	1.992**	2.110**	0.549	1.875**	2.242**	0.831	3.155*
MSE	0.029	0.021	0.035	0.023	0.037	0.022	0.035
Sharpe Ratio	0.012	0.063	-0.020	0.057	0.078	0.044	0.087
Ideal Profit	0.016	0.081	-0.026	0.074	0.102	0.057	0.113
	RNR-GARCH(1,1)		RNR-GJR GARCH				
<i>Sub-period</i>	Bull	Bear	Bull	Bear			
Total Return	0.177	0.623	0.303	0.606			
B&H Return	1.027	-0.982	1.027	-0.982			
Sign Rate	0.499	0.521	0.507	0.513			
PT test	-0.008	0.971	0.265	0.594			
HM test	-0.010	1.751**	0.344	1.104			
MSPE	0.022	0.046	0.021	0.034			
Sharpe Ratio	0.019	0.045	0.033	0.043			
Ideal Profit	0.025	0.058	0.043	0.056			

Notes: RNR = Recurrent Neural Network. Methods for forecasting volatility: MA(20) = Moving Average with a 20 days window, RM(0.94) = RiskMetrics’ exponentially weighted MA rule(decay factor = 0.94), GJR = Glosten-Jagannathan-Runkle (1993) GARCH model.

PT test = the Pesaran and Timmerman (1992) test. HT test = the Henriksson and Merton (1981) test. Both tests are asymptotically distributed as $N(0,1)$.

The sign rate measures the proportion of correctly predicted signs. The Sharpe ratio is defined as the ratio of the mean return of the strategy over its standard deviation. The Ideal Profit is the ratio of the returns of the trading strategy over the returns of a perfect predictor.

(*), (**), (***) indicate significance at the one sided 1%, 5% and 10% levels.

FIGURE 1: Daily closing prices of NASDAQ Index (02/08/1971 – 02/05/2002)



