

Non-linear Prediction of Security Returns with Moving Average Rules

RAMAZAN GENÇAY
University of Windsor, Canada

ABSTRACT

Over the years, investors and the technical analysts have devised hundreds of technical market indicators in an effort to forecast the trend of a security market. Recent literature provides evidence that these rules may provide positive profits after accounting for transaction costs. This clearly contradicts the theory of the efficient market hypothesis which states that security prices cannot be forecasted from their past values or *other past* variables. This paper uses the daily Dow Jones Industrial Average Index from January 1963 to June 1988 to examine the linear and non-linear predictability of stock market returns with buy-sell signals generated from the moving average rules with a band between the short and the long averages. Strong evidence of non-linear predictability is found in the stock market returns by using the past buy and sell signals of these rules.

KEY WORDS feedforward networks; technical trading rules; security returns

INTRODUCTION

Technical trading rule analysis assumes that there are patterns in market prices that will recur in the future and that these patterns can be used for predictive purposes. One popular rule for deciding when to buy and sell in a market is the moving average rule. This rule basically involves the calculation of a moving average of the raw price data. The simplest version of this rule indicates a buy signal whenever the price climbs above its moving average and a sell signal when it drops below. The underlying notion behind this rule is that it provides a means of determining the general direction or trend of a market by examining its recent history. For instance, an n -period moving average is computed by adding together the n most recent periods of data, then dividing by n . This average is recalculated each period by dropping the oldest data and adding the most recent, so the average *moves* with the data but does not fluctuate as much. An n -period moving average is *smoother* than an l -period (where $l < n$) moving average and measures a longer-term trend.

A typical moving average rule can be written as

$$m_t = (1/n) \sum_{i=0}^{n-1} p_{t-i} \quad (1)$$

Address for correspondence

CCC 0277-6693/96/030165-10
© 1996 by John Wiley & Sons, Ltd.

Received January 1995

According to equation (1) a buy signal is generated when the current price level p_t is above m_t ; otherwise a sell signal is generated. There are other variations of this rule. One is to replace the price series with a short moving average such that a buy signal is generated when the short moving average rises above the long moving average and a sell signal is generated otherwise. One of the most popular moving average rules used in technical analysis is the 1-200 rule, where the short period is one day and the long period is 200 days. Other popular ones are the 1-50, 1-150, 5-200 and the 2-200 rules. This paper tests the predictive power of the buy-sell signals generated from moving average rules with a band between the short and the long averages. The introduction of this band reduces the number of buy (sell) signals by eliminating *whiplash* signals when the short and long moving averages are close.

Contrary to technical trading analysis, the efficient market hypothesis states that security prices fully reflect all available information. A precondition for this strong version of the hypothesis is that information and trading costs are always zero. Since information and trading costs are positive, the strong form of the market efficiency hypothesis is clearly false. A weaker version of the efficiency hypothesis states that prices reflect information to the point where the marginal benefits of acting on information do not exceed the marginal costs (Jensen, 1978).

Earlier papers find evidence that daily, weekly and monthly returns are predictable from past returns. For example, Fama (1965) finds that the first-order autocorrelations of daily returns are positive for 23 of the 30 Dow Jones Industrials. Fisher's (1966) results suggest that the autocorrelations of monthly returns on diversified portfolios are positive and larger than those for individual stocks. As surveyed in Fama (1970, 1991), the evidence for predictability in these earlier works often lacks statistical power and the portion of the variance of returns explained by the variations in expected returns is so small that the hypothesis of market efficiency and constant expected returns is typically accepted as a good working model. More recent literature has investigated the predictability of current returns using other variables such as dividend yields and various term structure variables. This literature also documents significant relationships between expected returns and fundamental variables such as the price-earnings ratio, the market-to-book ratio and evidence of systematic patterns in stock returns related to various calendar periods such as the weekend, turn-of-the-month, holiday and January effects.

There has also been extensive recent work on the temporal dynamics of security returns. For instance, Lo and MacKinlay (1988) find that weekly returns on portfolios of NYSE stocks grouped according to size show positive autocorrelation. Conrad and Kaul (1988) examine the autocorrelations of Wednesday-to-Wednesday returns for size-grouped portfolios of stocks that trade on both Wednesdays. Similar to the findings of Lo and MacKinlay (1988), they find that weekly returns are positively autocorrelated. Cutler, Poterba and Summers (1991) present results from many different asset markets generally supporting the thesis that returns are positively correlated at the horizon of several months and negatively correlated at the 3-5 year horizon. Lo and MacKinlay (1990) report positive serial correlation in weekly returns for indices and portfolios and negative serial correlation for individual stocks. Chopra, Lakonishok and Ritter (1992), De Bondt and Thaler (1985), Fama and French (1988) and Poterba and Summers (1988) all find negative serial correlation in returns of individual stocks and various portfolios over three to ten year intervals. Jegadeesh (1990) finds negative serial correlation for lags up to two months and positive correlation for longer lags. Lehmann (1990) and French and Roll (1986) report negative serial correlation at the level of individual securities for weekly and daily returns. Overall, the findings of the recent literature confirm the earlier findings that daily and weekly returns are predictable from past returns and other economic and financial variables.

This evidence of the predictability of stock market returns has encouraged researchers to investigate the sources of this predictability. In Brock, Lakonishok and LeBaron (1992) (BLL hereafter), two of the simplest and most popular trading rules, moving averages and trading range breaks, are tested through the use of bootstrap techniques. They compare the returns conditional on buy (sell) signals from the actual Dow Jones Industrial Average (DJIA) Index to returns from simulated series generated from four popular null models. These null models are the random walk, the AR(1), the GARCH-M of Engle, Lilien and Robins (1987), and the exponential GARCH (EGARCH) developed by Nelson (1991). They find that returns obtained from buy (sell) signals are not likely to be generated by these four popular null models. They document that buy signals generate higher returns than sell signals and that the returns following buy signals are less volatile than returns after sell signals. In addition, they find that returns following sell signals are negative, which is not easily explained by any of the currently existing equilibrium models. Their findings indicate that the GARCH-M model fails not only in predicting returns but also in predicting volatility. They also document that the EGARCH model performs better than the GARCH-M in predicting volatility, although it also fails in matching the volatility during sell periods.

The results in BLL document two important stylized facts. The first is that buy signals consistently generate higher returns than sell signals. The second is that the second moments of the distributions of the buy and sell signals behave quite differently because returns following buy signals are less volatile than returns following sell signals. The asymmetric nature of the returns and the volatility of the Dow series over the periods of buy and sell signals suggest the existence of non-linearities. Overall, the findings of BLL show that linear conditional mean estimators fail to characterize the temporal dynamics of security returns and suggest the existence of possible non-linearities.

In a recent paper, Gençay (1995) investigates the predictability of security returns from buy-sell signals generated from the moving average rules where there is no band between the short and the long averages. Gençay's (1995) findings indicate that linear conditional mean specifications with past buy-sell signals as predictors of current returns provide slight forecast improvements over linear models of past returns. In non-linear conditional mean specifications, forecast improvements of the past buy-sell signals are significant over the linear conditional mean estimators with past returns as predictors of current returns. Overall, Gençay's (1995) results indicate that linear models of past buy-sell signals provide additional forecast improvement for current returns and the amount of forecastability is much more evident in non-linear conditional mean specifications.

This paper investigates the sensitivity of moving average rules to whiplash effects using a band between the short and long moving averages. The AR and GARCH-M models are used as linear conditional mean estimators. Feedforward networks are used as non-linear conditional mean estimators. As a measure of performance the out-of-sample mean square prediction error (MSPE) is used. The forecast horizon is specified to be 20 days. The results indicate that the adoption of the band eliminates noisy buy-sell signals and improves the quality of the out-of-sample forecasts. Moreover, the forecast performance of the non-linear conditional mean specifications get as large as 23% over the predictive performance of the linear benchmark model which uses past returns as predictors of current returns. In the next section a brief description of the data is presented. Estimation techniques are described in the third section and empirical results in the fourth. Conclusions follow thereafter.

DATA DESCRIPTION

The data set is the Dow Jones Industrial Average (DJIA) Index between 2 January 1963 and 30 June 1988, a total of 6409 observations. In addition to the whole data set, we also analyse four-year subsamples to measure the sensitivity of our results to sample variations.

Table I. Summary statistics of the log first differenced daily DJIA series January 1963–June 1988

Description	1963–88	1963–67	1968–71	1972–75	1976–19	1980–83	1984–88
Sample size	6409	1258	982	1008	1009	1011	1136
Mean*100	0.0187	0.0267	-0.0019	-0.0042	-0.0023	0.0418	0.0472
Std.*100	0.9598	0.5780	0.7503	1.0960	0.7709	0.9775	1.3752
Skewness	-2.8059	0.0589	0.4932	0.2091	0.1650	0.3592	-5.7253
Kurtosis	86.8674	7.1234	6.3526	3.9791	4.1694	4.3425	113.2671
Max	0.0967	0.0440	0.0495	0.0460	0.0436	0.0478	0.0967
Min	-0.2563	-0.0293	-0.0319	-0.0357	-0.0304	-0.0359	-0.2563
ρ_1	0.1036	0.1212	0.2929	0.2118	0.1130	0.0470	0.0126
ρ_2	-0.0390	0.0269	-0.0024	-0.0531	0.0090	0.0480	-0.1051
ρ_3	-0.0083	0.0243	0.0046	-0.0099	0.0197	-0.0228	-0.0172
ρ_4	-0.0231	0.0480	0.0485	-0.0260	-0.0188	-0.0361	-0.0466
ρ_5	0.0247	0.0267	0.0248	-0.0627	-0.0051	-0.0243	0.1015
ρ_6	-0.0098	0.0153	-0.0639	-0.0360	-0.0530	0.0293	0.0058
ρ_7	0.0065	0.0014	-0.0314	0.0072	0.0114	-0.0130	0.0234
ρ_8	-0.0029	0.0396	0.1087	-0.0054	-0.0632	-0.0164	-0.0151
ρ_9	-0.0133	0.0107	0.0086	-0.0487	0.0216	0.0099	-0.0230
ρ_{10}	-0.0133	-0.0064	-0.0619	-0.0048	0.0166	-0.0205	-0.0131
Bartlett std.	0.0125	0.0282	0.0319	0.0315	0.0315	0.0314	0.0297
LBP	89.6	25.3	109.0	57.3	21.8	8.96	29.5
$\chi^2_{0.05}(10)$	8.307						

Notes:

ρ_1, \dots, ρ_{10} are the first ten autocorrelations of each series. LBP refers to the Ljung–Box–Pierce statistic and it is distributed $\chi^2(10)$ under the null hypothesis of identical and independent observations.

The summary statistics of the daily returns are presented in Table I. The daily returns are calculated as the log differences of the Dow level. The 1984–88 subperiod is highly skewed and exhibits extremely high kurtosis. The influence of this subperiod is reflected in the skewness and kurtosis values of the whole data set. The other subperiods do not exhibit statistically significant skewness values but there is evidence of kurtosis in these subperiods.

The first 10 autocorrelations are labelled ρ_n and are reported together with the Bartlett standard errors. The Ljung–Box–Pierce statistics are shown in the last row. These are calculated for the first 10 lags and are distributed $\chi^2(10)$ under the null of identical and independently distributed returns. All periods except 1980–83 show evidence of autocorrelation.

ESTIMATION TECHNIQUES

Let $p_t, t = 1, 2, \dots, T$, be the daily Dow series. The return series is calculated as $r_t = \log(p_t) - \log(p_{t-1})$. Let m_t^n denote the time t value of a moving average rule of length n . m_t^n is calculated by

$$m_t^n = (1/n) \sum_{i=0}^{n-1} p_{t-i} \quad (2)$$

The buy and sell signals are calculated by

$$s_i^{n1, n2} = \begin{cases} (m_i^{n1} - m_i^{n2}) & \text{if } |m_i^{n1} - m_i^{n2}| > \text{band} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where $n1$ and $n2$ are the short and the long moving averages, respectively. The rules used in this paper are $(n1, n2) = [(1, 50), (1, 200), (5, 150), (2, 200)]$ where $n1$ and $n2$ are in days. The band is set to 0.01 for all the rules studied above. This is a small band that eliminates whiplash effects from the data.

The linear test regression is

$$r_t = \alpha + \sum_{i=1}^p \beta_i s_{t-i}^{n1, n2} + \varepsilon_t \quad (4)$$

where $\varepsilon_t \sim \text{ID}(0, \sigma_t^2)$. In the case of the GARCH-M(1,1) process the test model is written as

$$r_t = \alpha + \sum_{i=1}^p \beta_i s_{t-i}^{n1, n2} + \gamma h_t + \varepsilon_t \quad (5)$$

where $\varepsilon_t = h_t u_t$, $u_t \sim \text{NID}(0, 1)$ and $h_t^2 = \delta_0 + \delta_1 h_{t-1}^2 + \delta_2 \varepsilon_{t-1}^2$.

There are numerous nonparametric regression techniques available, such as flexible Fourier forms, non-parametric kernel regression, wavelets, spline techniques and artificial neural networks. Here, a class of artificial neural network models, namely the single-layer feedforward networks, are used. The justification for this choice is that the rate of convergence of these networks does not depend on the dimensionality of the input space. Recently, Hornik *et al.* (1994) have shown that single hidden-layer feedforward networks can approximate unknown functions and their derivatives with error decreasing at rates as fast as $d^{-1/2}$ (d is the number of hidden units in a network) and that the dimension of the input space, p , does not affect the rate of approximation, but only the constants of proportionality. This is in sharp contrast to the properties of the standard kernel and series approximants. This is an advantage in terms of having desirable estimators in small samples. The single-layer feedforward network regression model with lagged buy and sell signals and with d hidden units is written as

$$r_t = \alpha_0 + \sum_{j=1}^d \beta_j G\left(\alpha_j + \sum_{i=1}^p \gamma_{ij} s_{t-i}^{n1, n2}\right) + \varepsilon_t, \quad \varepsilon_t \sim \text{ID}(0, \sigma_t^2) \quad (6)$$

where G is a known activation function. Here, G is chosen to be

$$G(x) = \frac{1}{1 + \exp(-\gamma x)}$$

the logistic function which is a sigmoidal.¹ Sigmoidal functions satisfy the universal approximation results for unknown functions derived by Hornik, Stinchcombe and White (1989, 1990). Here, only the logistic function is used, this being widely employed in the artificial neural network literature.

Many authors have investigated the universal approximation properties of neural networks (Gallant and White, 1988, 1992; Cybenko, 1989; Funahashi, 1989; Hecht-Nielson, 1989;

¹ G is a sigmoid function if $G: \mathbb{R} \rightarrow [0, 1]$, $G(x) \rightarrow 0$ as $x \rightarrow -\infty$, $G(x) \rightarrow 1$ as $x \rightarrow \infty$ and G is monotonic.

Hornik, Stinchcombe and White (1989, 1990). Using a wide variety of proof strategies, all have demonstrated that, under general regularity conditions, a sufficiently complex single hidden-layer feedforward network can approximate any member of a class of functions to any desired degree of accuracy, where the complexity of a single hidden-layer feedforward network is measured by the number of units in the hidden layer. For an excellent survey of feedforward and recurrent network models, the reader may refer to Kuan and White (1994).

To compare the performance of the regression models in equations (4)–(6) the linear regression

$$r_t = \alpha + \sum_{i=1}^p \beta_i r_{t-i} + \varepsilon_t, \quad \varepsilon_t \sim \text{ID}(0, \sigma_t^2) \quad (7)$$

is used as the benchmark model. The out-of-sample forecast performance of equations (4)–(6) are measured by the ratio of their mean square prediction error's (MSPE) to that of the linear benchmark model in equation (7).

A number of papers in the literature suggest that conditional heteroscedasticity may be important in the improvement of the forecast performance of the conditional mean. For this reason, the MSPE of the GARCH-M(1,1) model with lagged returns

$$r_t = \alpha + \sum_{i=1}^p \beta_i r_{t-i} + \gamma h_t + \varepsilon_t \quad (8)$$

where $\varepsilon_t = h_t u_t$, $u_t \sim \text{NID}(0, 1)$ and $h_t^2 = \delta_0 + \delta_1 h_{t-1}^2$ is compared to that of the benchmark model in equation (7).

The out-of-sample forecast performance of the single-layer feedforward network model with lagged returns

$$r_t = \alpha_0 + \sum_{j=1}^d \beta_j G \left(\alpha_j + \sum_{i=1}^p \gamma_{ij} r_{t-i} \right) + \varepsilon_t, \quad \varepsilon_t \sim \text{ID}(0, \sigma_t^2) \quad (9)$$

is also compared to that of the benchmark model in equation (7).

EMPIRICAL RESULTS

Any evidence of predictability should result from a careful analysis of out-of-sample forecasts which are designed so that problems stemming from spurious results relating to sample specific conditions are avoided. This section provides a detailed out-of-sample analysis of 25 years of daily data split into six subsamples. The advantage of studying the forecasts from multiple subsamples is that they enable us to analyse the performance of technical trading rules under different market conditions: it is especially important to observe the performance of these rules in trendy versus sluggish market conditions in which there is no clear trend in either direction.

For each subsample, the out-of-sample predictive performance of the benchmark and test models are examined. The out-of-sample prediction is done by excluding the last 20 observations from in-sample estimation. The out-of-sample forecasts with the buy-sell signals as the conditioning set are computed recursively by estimating $E(r_t | s_{t-1}^{n1, n2}, \dots, s_{t-p}^{n1, n2})$, then $E(r_{t+1} | s_{t-1}^{n1, n2}, \dots, s_{t-p+1}^{n1, n2})$, and so forth in real time. This is continued until the sample is exhausted, resulting in a sequence of 20 *ex ante* one-step-ahead forecasts. Similarly, the out-of-sample forecasts with past returns as the conditioning set are computed recursively by

estimating $E(r_t | t_{r-1}, \dots, r_{t-p})$ and then $E(r_{t+1} | r_t, \dots, r_{t-p+1})$ until the sample is exhausted for each forecast horizon.

Let MSPE' and MPSE^b be the mean square prediction errors of the test and benchmark models, respectively. To measure the out-of-sample performance between the test and benchmark models, the ratio of the mean square prediction errors, MSPE'/MSPE^b is used. MSPE'/MSPE^b is less than one if the test model provides more accurate predictions. Similarly, the ratio is greater than one if the predictions of the test model are less accurate relative to the benchmark model.

For all models, up to four lags of returns are used to predict current stock returns. One of several different ways of measuring the trade off between the goodness of fit and parsimony is the Schwartz Information Criterion

$$\text{SIC} = \log(\text{MSE}) + k[\log(T)]/T \quad (10)$$

where k is the number of parameters in the conditional mean and T is the number of observations. As indicated in Holmes and Hutton (1989) the SIC performs well in Monte Carlo simulation studies.

Each feedforward network model is estimated with up to 10 hidden units for each data set. For each data set, there are four models corresponding to four different lags for the linear models with past returns. For feedforward network models with past returns there are 40 models corresponding to four lags and 10 hidden network models. For the linear models with past buy-sell signals, there are 20 models corresponding to four lags and five rules. For the feedforward network models there are 200 choices corresponding to four lags, five rules and 10 hidden network models. For each sample period, the SIC is used as the model selection criterion.

Empirical results with past returns

In this study, a 10% forecast improvement is adopted as a benchmark measure of performance. The ratio of mean square prediction errors which provide a 10% or more forecast improvement are underlined in each table.

In Table II, the ratio of the MSPE of the corresponding model to the MSPE of the benchmark model are reported. For all subperiods and the whole data set, the forecast performance of the GARCH-M (1,1) model is equivalent to that of the benchmark model. The last column of Table II reports the results with the feedforward network. In the two subsamples, namely the 1968-71 and 1972-75 periods, the feedforward network achieves the 10% forecast performance over the benchmark model. The forecast performance of the feedforward network model in the other periods is comparable to that of the benchmark model.

Table II. MSPEs of the models with past returns

Date	Obs.	Lags	Benchmark	Lags	GARCH-M (1,1)	Lags	Feedforward
2/1/1963-29/12/1967	1259	1	0.1548	1	0.96	3	0.96
2/1/1968-31/12/1971	983	1	0.3910	1	1.00	2	<u>0.88</u>
3/1/1972-31/12/1975	1009	1	0.6298	1	0.99	2	<u>0.80</u>
2/1/1976-31/12/1979	1010	2	0.1654	2	0.98	4	<u>0.97</u>
2/1/1980-30/12/1983	1012	2	0.2998	2	0.99	3	1.00
3/1/1984-30/06/1988	1137	1	1.1853	1	1.02	4	0.94
2/1/1963-30/06/1988	6409	1	1.2700	1	1.03	4	1.04

Notes:

MSPEs of the benchmark model are reported in levels. The MSPEs of other models are reported as a ratio to that of the benchmark model. Lags report the lags of the corresponding model indicated by the SIC.

Empirical results with past buy-sell signals of the moving average rules

Each of the OLS, GARCH-M (1,1) and the feedforward network models are estimated with up to four lags of the buy-sell signals as regressors. The four moving average rules described earlier are utilized in this section. These rules are $(n1, n2) = [(1, 50), (1, 200), (5, 150), (2, 200)]$ where $n1$ and $n2$ are in days. For convenience, these rules are referred to as $a = (1, 50)$, $b = (1, 200)$, $c = (5, 150)$ and $d = (2, 200)$ in the following tables. In addition to the four rules above we also estimate a combined rule which we refer to as rule v . The specification of the feedforward network model with this rule is

$$r_t = \alpha_0 + \sum_{j=1}^d \beta_j G(\alpha_j + \gamma_{1j} s_{t-1}^{1,50} + \gamma_{2j} s_{t-1}^{1,200} + \gamma_{3j} s_{t-1}^{2,200} + \gamma_{4j} s_{t-1}^{5,150}) + \varepsilon_t \quad (11)$$

For the OLS and the GARCH-M models, the regressors $s_{t-1}^{1,50}$, $s_{t-1}^{1,200}$, $s_{t-1}^{2,200}$ and $s_{t-1}^{5,150}$ enter linearly in the conditional mean specifications.

In Table III, the comparison of the forecast performances of the moving average rules with and without a band are presented under columns represented by % Δ . For the OLS model, the results indicate that the moving average rule with $band = 0.01$ performs 1-2% better than the rule without a band for all forecast horizons. The performance of the GARCH-M (1,1) model is comparable to the performance of the OLS model in Table III. The forecast performance of the GARCH-M (1,1) models is 1-2% in favour of the moving average rules with a band. The results in Table III indicate that feedforward networks provide substantial forecast improvements over the benchmark model. Feedforward networks consistently achieve the 10% forecast gain over the benchmark model for all data periods and the largest forecast gain is as much as 23% for the 1984-88 period. Forecast performance of the moving average rules with a band remains 1-2% better than the rules without a band for most of the periods and this differential goes up to 5% in the 1976-79 period.

There are two results which emerge from the feedforward network estimation with past buy-sell signals. The first is that the introduction of the band around a moving average rule improves forecast accuracy. The second result is that there are substantial forecast gains from feedforward network estimation with past buy-sell signals. This may not necessarily imply profitability of these rules as the transaction costs and brokerage fees are not taken into account.

Table III. MSPEs of the models with past buy-sell signals

Forecast horizon	Date	Obs.	OLS			GARCH-M			FN		
			Lags	Ratio	% Δ	Lags	Ratio	% Δ	Lags	Ratio	% Δ
	2/1/1963-29/12/1967	1259	b3	0.94	2	a3	0.91	1	v1	<u>0.86</u>	3
	2/1/1968-31/12/1971	983	c3	0.93	1	c3	0.95	2	b2	<u>0.78</u>	2
	3/1/1972-31/12/1975	1009	d4	0.95	1	d4	0.96	0	c2	<u>0.82</u>	0
	2/1/1976-31/12/1979	1010	v4	0.98	1	b2	0.98	0	d2	<u>0.80</u>	5
	2/1/1980-30/12/1983	1012	b4	0.94	1	c3	0.94	1	c2	0.91	2
	3/1/1984-30/06/1988	1137	v2	0.94	1	b1	0.97	1	c3	<u>0.77</u>	2
	2/1/1963-30/06/1988	6409	c2	0.91	2	v2	0.97	0	c1	<u>0.84</u>	2

Notes:

The column entitled *Lags* reports the rule and the lags of the corresponding model indicated by the SIC. *Ratio* refers to the ratio of the MSPE of the corresponding model to the MSPE of the benchmark model. % Δ refers to the percentage forecast improvement between models with and without a band. GARCH-M refers to the GARCH-M (1,1) estimation and FN refers to the feedforward networks.

In fact, given the large number of potential transactions that these rules suggest, it may be the case that if the latter are accounted for then the predictive power of these rules may be compromised by their lack of profitability.

CONCLUSIONS

This paper has investigated the out-of-sample predictive performance of the moving average rules with a band in linear and non-linear conditional mean specifications. The results indicate that non-linear conditional mean specifications of the past buy-sell signals of the moving average rule with a band provide forecast improvements for the current returns over the linear benchmark model, which uses past returns as regressors. These rules consistently provide better out-of-sample predictions than the moving average rules without a band, although the difference between the out-of-sample predictions of these two types of rules are limited to an average of 1–2%. The accuracy of the out-of-sample predictions of these rules is much more evident in the non-linear conditional mean specifications.

ACKNOWLEDGEMENTS

I thank the editor; an anonymous referee for detailed comments; Buz Brock and Blake LeBaron for discussions. I also thank Blake LeBaron for providing the data; Chung-Ming Kuan and Tung Liu for providing some of the computing algorithms; and the Natural Sciences and Engineering Research Council of Canada and the Social Sciences and Humanities Research Council of Canada for financial support.

REFERENCES

- Brock, W. A., Lakonishok, J. and LeBaron, B., 'Simple technical trading rules and the stochastic properties of stock returns', *Journal of Finance*, 47 (1992), 1731–64.
- Chopra, N., Lakonishok, J. and Ritter, J., 'Performance measurement methodology and the question of whether stocks overreact', *Journal of Financial Economics*, 31 (1992), 235–68.
- Conrad, J. and Kaul, G., 'Time-variation in expected returns', *Journal of Business*, 61 (1988), 409–25.
- Cutler, D. M., Poterba, J. M. and Summers, L. H., 'Speculative dynamics', *Review of Economic Studies*, 58 (1991), 529–546.
- Cybenko, G., 'Approximation by superposition of a sigmoidal function', *Mathematics of Control, Signals and Systems*, 2 (1989), 303–14.
- De Bondt, W. F. M. and Thaler, R. H., 'Does the stock market overreact?' *Journal of Finance*, 40 (1985), 793–805.
- Engle, R. F., Lilien, D. and Robins, R. P., 'Estimating time varying risk premia in the term structure: The ARCH-M model', *Econometrica*, 55 (1987), 391–407.
- Fama, E. F., 'The behavior of stock market prices', *Journal of Business*, 38 (1965), 34–105.
- Fama, E. F., 'Efficient capital markets: A review of theory and empirical work', *Journal of Finance*, 25 (1970), 383–417.
- Fama, E. F., 'Efficient capital markets: II', *Journal of Finance*, 46 (1991), 1575–1617.
- Fama, E. F. and French, K. R., 'Permanent and temporary components of stock prices', *Journal of Political Economy*, 96 (1988), 246–273.
- Fisher, L., 'Some new stock-market indexes', *Journal of Business*, 39 (1966), 191–225.
- French, K. R. and Roll, R., 'Stock return variances: The arrival of information and the reaction of traders', *Journal of Financial Economics*, 17 (1986), 5–26.
- Funahashi, K.-I., 'On the approximate realization of continuous mappings by neural networks', *Neural Networks*, 2 (1989), 183–192.

- Gallant, A. R. and White, H., 'There exists a neural network that does not make avoidable mistakes', *Proceedings of the Second Annual IEEE Conference on Neural Networks*, San Diego, CA, I.657-I.664, New York: IEEE Press.
- Gallant, A. R. and White, H., 'On learning the derivatives of an unknown mapping with multilayer feedforward networks', *Neural Networks*, 5 (1992), 129-138.
- Gençay, R., 'The predictability of security returns with simple technical trading rules', University of Windsor, unpublished manuscript, 1995.
- Hecht-Nielsen, R., 'Theory of the backpropagation neural networks', *Proceedings of the International Joint Conference on Neural Networks*, Washington, DC, I.593-I.605, New York IEEE Press.
- Holmes, J. M. and Hutton, P. A., 'Optimal' model selection when the true relationship is weak and occurs with a delay', *Economics Letters*, 30 (1989), 333-9.
- Hornik, K., Stinchcombe, M. and White, M., 'Multilayer feedforward networks are universal approximators', *Neural Networks*, 2 (1989), 359-366.
- Hornik, K., Stinchcombe, M. and White, H., 'Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks', *Neural Networks*, 3 (1990), 551-60.
- Hornik, K., Stinchcombe, M., White, H. and Auer, P., 'Degree of approximation results for feedforward networks approximating unknown mappings and their derivatives', UCSD discussion paper, 1994.
- Jegadeesh, N., 'Evidence of predictable behavior of security returns', *Journal of Finance*, 45 (1990), 881-898.
- Jensen, M. C., 'Some anomalous evidence regarding market efficiency', *Journal of Financial Economics*, 6 (1978), 95-101.
- Kuan, C.-M. and White, H., 'Artificial neural networks: An econometric perspective', *Econometric Reviews*, 13, (1994), 1-91.
- Lehmann, B. N., 'Fads, martingales and market efficiency', *Quarterly Journal of Economics*, 105, (1990), 1-28.
- Lo, A. W. and MacKinlay, A. C., 'Stock market prices do not follow random walks: Evidence from a simple specification test', *Review of Financial Studies*, 1 (1988), 41-66.
- Lo, A. W. and MacKinlay, A. C., 'When are contrarian profits due to stock market overreaction?', *Reviews of Financial Studies*, 3 (1990), 175-205.
- Nelson, D. B., 'Conditional heteroskedasticity in asset returns: A new approach', *Econometrica*, 59 (1991), 347-370.
- Poterba, J. M. and Summers, L. H., 'Mean reversion in stock prices: Evidence and implications', *Journal of Financial Economics*, 22 (1988), 27-59.

Author's biography:

Ramazan Gençay is an Associate Professor in the Economics Department at the University of Windsor. His areas of specialization are the detection of chaotic dynamics from observed data, nonlinear time series modelling and financial forecasting. Some of his publications have appeared in the *Journal of the American Statistical Association*, *Physica D*, *Journal of Nonparametric Statistics* and the *Journal of Applied Econometrics*.

Author's address:

Ramazan Gençay, Department of Economics, University of Windsor, 401 Sunset, Windsor, Ontario N9B 3P4, Canada.