Studies in Nonlinear Dynamics and Econometrics

Quarterly Journal
July 1997, Volume 2, Number 2
The MIT Press

Studies in Nonlinear Dynamics and Econometrics (ISSN 1081-1826) is a quarterly journal published electronically on the Internet by The MIT Press, Cambridge, Massachusetts, 02142. Subscriptions and address changes should be addressed to MIT Press Journals, Five Cambridge Center, Cambridge, MA 02142; tel.: (617) 253-2889; fax: (617) 577-1545; e-mail: journals-orders@mit.edu. Subscription rates are: Individuals $40.00, Institutions $130.00. Canadians add additional 7% GST. Prices subject to change without notice.

Subscribers are licensed to use journal articles in a variety of ways, limited only as required to insure fair attribution to authors and the Journal, and to prohibit use in a competing commercial product. See the Journal's World Wide Web site for further details. Address inquiries to the Subsidiary Rights Manager, MIT Press Journals, Five Cambridge Center, Cambridge, MA 02142; tel.: (617) 253-2864; fax: (617) 258-5028; e-mail: journals-rights@mit.edu.

© 1997 by the Massachusetts Institute of Technology

Technical Trading Rules and the Size of the Risk Premium in Security Returns

Ramazan Gençay
Department of Economics
University of Windsor
gencay@uwindsor.ca

Thanasis Stengos
Department of Economics
University of Guelph

Abstract. *Among analysts, technical trading rules are widely used for forecasting security returns. Recent literature provides evidence that these rules may provide positive profits after accounting for transaction costs. This would be contrary to 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 1963 to 1988 to examine the linear and nonlinear predictability of stock market returns with simple technical trading rules, by using the nearest neighbors and the feedforward network regressions. Evidence of nonlinear predictability is found in the stock market returns by using the past returns and the buy and sell signals of the moving average rules.*

Keywords. market efficiency, technical trading rules, feedforward networks

Acknowledgments. We would like to thank the helpful comments of an associate editor and two referees. Both authors would like to thank the Social Sciences and Humanities Research Council of Canada for financial support. Ramazan Gençay also thanks the Natural Sciences and Engineering Research Council of Canada for financial support.

1 Introduction

Technical analysts test historical data to establish specific rules for buying and selling securities with the objective of maximizing profit and minimizing risk of loss. Technical trading analysis is based on two main premises. First, the market's behavior patterns do not change much over time, particularly the longer-term trends. While future events can indeed be very different from any past events, the market's way of responding to brand-new uncertainties is usually similar to the way it handled them in the past. The patterns in market prices are assumed to recur in the future, and thus, these patterns can be used for predictive purposes. Second, relevant investment information may be distributed fairly efficiently, but it is not distributed perfectly, nor will it ever be. Even if it were, some investors, through superior analysis and insight, would always have

© 1997 by the Massachusetts Institute of Technology

an edge over the majority of investors and would act first. Therefore, valuable information can be deduced by studying transaction activity.

One common component of many technical rules 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 whenever 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 the 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 its data, but does not fluctuate as much.

A typical moving average rule can be written as

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

According to Equation (1), a buy signal is generated when the current price level p_t is above m_t, $(p_t - m_t) > 0$; otherwise a sell signal is generated. The most popular moving-average rule as reported in Brock, Lakonishok, and LeBaron (1992) 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.

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 not very useful empirically. 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 works 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 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.

Unlike the earlier literature which focused on the predictability of current returns from past returns, the recent literature has also investigated the predictability of current returns from 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 for systematic patterns in stock returns related to various calendar periods such as the weekend effect, the turn-of-the-month effect, the holiday effect, and the January effect.

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 (to mitigate the nonsynchronous trading problem) 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 hypothesis 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 (1986), and Poterba and Summers (1988) 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 recent literature confirm the findings of earlier literature that the daily and weekly returns are predictable from past returns and other economic and financial variables.

Evidence of the inefficiency of stock market returns led the researchers to investigate the sources of this inefficiency. In Brock, Lakonishok, and LeBaron (1992) (BLL hereafter), two of the simplest and most popular trading rules, moving-average and the trading-range brake rules, 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 due to 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 the returns following buy signals are less volatile than returns on 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 distribution of the buy and sell signals behave quite differently because the 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 nonlinearities as the data-generation mechanism. Overall, the findings of BLL show that the linear conditional mean estimators fail to characterize the temporal dynamics of the security returns and suggest the existence of possible nonlinearities.

This paper uses the rule in Equation (1) to investigate the predictive power of simple technical trading rules in forecasting the current returns. The test regressions of this paper contain the past buy and sell signals of the technical trading rules in Equation (1) as regressors to forecast the current returns. To measure the performance of the test regressions, the random walk model is studied as the benchmark model. The simple GARCH-M (1, 1) model is used as the linear parametric conditional mean estimator. We use nonparametric regressions to capture any possible nonlinearities in the conditional means. In this study, we compare the out-of-sample performances of two nonparametric conditional mean estimators to those of the parametric ones. These nonparametric regression models are the feedforward network regression and the nearest neighbors regression models. The first one is a global estimator, whereas the second is a local procedure.

As a measure of performance, we use the out-of-sample root mean square prediction error (RMSPE). The statistical significance of the RMSPEs of the test models against the random walk model are studied by the Mizrach (1995) test. The data set is the daily DJIA Index from January 1, 1963 to June 30, 1988, studied in six subsamples. For each subsample, the forecast horizon is chosen to be 250 observations, a year of price data in daily frequency. There are two advantages of constructing the forecast horizon from six different subsamples. The first one is to avoid spurious results caused by data snooping problems or sample-specific conditions. The second one is that it enables us to analyze the performance of the technical trading rules under different market conditions. This is particularly important in observing the performance of these rules in trendy versus sluggish market conditions in which there is no clear trend in either direction.

The results of this paper indicate that in general the nonparametric regression forecasts provide improvements over the forecasts of the parametric model that we considered in terms of reduced RMSPE,

Table 1
Summary Statistics of the Log First Differenced Daily DJIA Series: January 1963–June 1988

Description	1963–1988	1963–1967	1968–1971	1972–1975	1976–1979	1980–1983	1984–1988
Sample Size	**6,404**	**1,258**	**982**	**1,008**	**1,009**	**1,011**	**1,136**
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
Maximum	0.0967	0.0440	0.0495	0.0460	0.0436	0.0478	0.0967
Minimum	−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
$\chi^2_{0.05}(10)$	18.307						

Notes: $\rho_1, \ldots, \rho_{10}$ are the first 10 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.

both for the trading rules we considered as well as for an even simpler model based on past returns. Among the nonparametric models, the forecasts of the local procedure (nearest neighbors regression) offers more gains than the forecasts of the global procedure (feedforward regression) when measured against the benchmark random walk model. The nonparametric models that use past buy-sell signals provide better RMSPEs relative to the models that use past returns. The RMSPE test of Mizrach (1995) is a test for comparing predictors of univariate time series in the mean squared error. We use this test to compare the forecasts of the random walk model against the parametric and the nonparametric model forecasts. Mizrach (1995) tests indicate that RMSPEs of the nonparametric models are statistically significant against the RMSPEs of the random walk model.

In Section 2, a brief description of the data is presented. Estimation techniques are described in Section 3, and empirical results are in Section 4. Conclusions follow thereafter.

2 Data Description

The data series includes the first trading day in 1963 of the Dow Jones Industrial Average Index to June 30, 1988, a total of 6,404 observations. All of the stocks are actively traded and problems associated with nonsynchronous trading should be of little concern with the DJIA.

The data set is studied in six subsamples. The summary statistics of the daily returns for all subsamples are presented in Table 1. The daily returns are calculated as the log differences of the Dow level. None of the subperiods except the 1983–1988 period show significant skewness, and the majority of the subperiods exhibit excess kurtosis.

The first 10 autocorrelations are also given in the rows labelled ρ_n. The Bartlett standard errors from these series are also reported in Table 1. All periods show some evidence of autocorrelation in the first lag, except the 1983–1988 period. 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 independent observations. In five subperiods out of six, the null hypothesis of identical and independent observations is rejected.

3 Estimation Techniques

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

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

The buy and sell signals are calculated by

$$s_t^{n1,n2} = m_t^{n1} - m_t^{n2} \tag{3}$$

where $n1$ and $n2$ are the short- and long-moving averages, respectively. The rules used in this paper are $(n1, n2) = [(1, 50), (1, 200)]$, where $n1$ and $n2$ are in days. The GARCH-M(1,1) is used as the parametric test model and is written as

$$r_t = \alpha + \sum_{i=1}^{p} \beta_i s_{t-i}^{n1,n2} + \gamma h_t^{1/2} + \epsilon_t \tag{4}$$

where $\epsilon_t \sim N(0, h_t)$ and $h_t = \delta_0 + \delta_1 h_{t-1} + \delta_2 \epsilon_{t-1}^2$. The GARCH-M specification allows for the conditional second moments of the return process to be serially correlated. This specification implies that periods of high (low) volatility are likely to be followed by periods of high (low) volatility. The GARCH-M specification allows for the volatility to change over time and the expected returns are a function of past returns as well as volatility.

There are numerous nonparametric regression techniques available, such as flexible Fourier forms, nearest neighbors regression, nonparametric kernel regression, wavelets, spline techniques, and artificial neural networks. Here, a class of artificial neural network models, namely the single-layer feedforward networks and nearest neighbors regression models, are used. These two nonparametric regression models are described below.

3.1 Nearest neighbors regression

The conditional mean of a random variable x, given a vector of conditioning variables w, can be written as $E(x|w) = M(w)$. In parametric estimation, $M(w)$ is typically assumed to be linear in w, but in the nonparametric approach, $M(w)$ remains a general functional form. In this paper, we take a simple approach to forecasting $M(w)$, using the nearest neighbor method of Stone (1977). Applications of nearest neighbor methods include the work of Robinson (1987) in a regression context, as well as the work of Yakowitz (1987) in a time-series forecasting context.

Consider now the time-series process $\{x_t\}$, and in particular, the problem of estimating the mean of x_t conditional on $(x_{t-1}, \ldots, x_{t-n})$. The nearest neighbor method can be intuitively explained in the following way. Take the time series $\{x_t\}_{t=1}^{T}$and convert it into a series of vectors of n components each, denoted as $x_t^n = (x_t, x_{t-1}, \ldots, x_{t-n+1})$. The above n vectors represent n past histories of the process $\{x_t\}$. Now for the nearest neighbor forecasting problem, one takes the most recent history available and searches over the set of all n histories to find the k nearest neighbors. For instance, if one wants to forecast x_t from the information available at $t-1$, one computes the distance of the vector x_{t-1}^n defined as $x_{t-1}^n = (x_{t-1}, x_{t-2}, \ldots, x_{t-n})$ and its k nearest neighbors to form an alternative estimator of $E(x_t|x_{t-1}, x_{t-2}, \ldots, x_{t-n})$ by $\sum_{i=1}^{k} \omega_{ti} x_i$, where ω_{ti} represents the k nearest neighbor weight. Typically, one uses the Euclidean distance to compute these weights. For a more thorough discussion of the weighting schemes that are available for the construction of the above distance, see Robinson (1987). From a computational point of view, uniform weights are the most popular in the literature; see Härdle (1990). Also, the choice of weights will only affect the bias and the variance contribution terms to the mean square error up to a proportionality factor. Hence, asymptotically, the choice of weights is not important, although there may be small sample effects. In the present application, we used uniform weights to weigh the contribution of the k nearest neighbors in the overall estimate of the regression function $E(x_t|x_{t-1}, x_{t-2}, \ldots, x_{t-n})$. The choice of nearest neighbors was determined by cross-validation; see Härdle (1990). In the context of the present application, the estimator of the regression function is given as

$$\hat{E}(r_t|s_{t-1}^{n1,n2}, s_{t-2}^{n1,n2}, \ldots, s_{t-p}^{n1,n2}) = \sum_{i=1}^{k} \omega_{ti} r_i. \tag{5}$$

In using Equation (5) above, for $p = 2$, the k nearest neighbors are chosen to be the ones closest to $(s_{t-1}^{n1,n2}, s_{t-2}^{n1,n2})$ from the set of $(s_{t-1}^{n1,n2}, s_{t-2}^{n1,n2})$, $t = 2, \ldots, n$. Then, the estimator of $E(r_t|s_{t-1}^{n1,n2}, s_{t-2}^{n1,n2})$ is computed as a simple average of the r_s terms, corresponding to the k nearest neighbors.

For each one-step-ahead forecast observation, the nearest neighbors regression is re-estimated and the optimal nearest neighbors regression complexity is determined according to the one-leave-out cross-validated performance measure. Accordingly, a different model may be indicated by the cross-validated performance measure at different forecast horizons. A rolling sample approach is used, so that the same number of observations are used as the in-sample observations at every one-step-ahead prediction. The maximum number of nearest neighbors ($k = \{1, 2, \ldots, 100\}$) and the maximum number of lags ($p = 5$) is chosen according to the computational limitations.

3.2 Feedforward networks

The single-layer feedforward network regression model with past buy and sell signals and with d hidden units is written as

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

where G is the known activation function that is chosen to be the logistic function. This choice is common in the artificial neural networks literature. Many authors have investigated the universal approximation properties of neural networks (Gallant and White 1988, 1992; Cybenko 1989; Funahashi 1989; Hecht-Nielson 1989; 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 hidden units in the hidden layer. For an excellent survey of the 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), (5), and (6), the random walk

$$r_t = \alpha + \epsilon_t \qquad \epsilon_t \sim ID(0, \sigma^2) \tag{7}$$

is used with the lagged returns as the benchmark model. The out-of-sample forecast performance of Equations (4), (5), and (6) is measured by the ratio of their RMSPEs to that of the linear benchmark model in Equation (7). A number of papers in the literature suggest that conditional heteroskedasticity may be important in the improvement of the forecast performance of the conditional mean. For this reason, the RMSPE of the GARCH-M (1,1) model with lagged returns

$$r_t = \alpha + \sum_{i=1}^{p} \beta_i r_{t-i} + \gamma h_t^{1/2} + \epsilon_t \quad \epsilon_t \sim N(0, h_t) \quad h_t = \delta_0 + \delta_1 h_{t-1} + \delta_2 \epsilon_{t-1}^2 \tag{8}$$

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

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

is also compared to that of the benchmark model in Equation (7). Feedforward network regression models require a choice for the number of hidden units in a network. Let

$$o_t = \alpha_0 + \sum_{j=1}^{d} \beta_j G\left(\alpha_j + \sum_{i=1}^{p} \gamma_{ij} x_{t-i}\right) \tag{10}$$

where x_{t-i} is either past returns (Equation 9) or past buy-sell signals (Equation 6). The cross-validated performance measure is formally defined as

$$C_T(d) \equiv T^{-1} \sum_{t=1}^{T} [r_t - \hat{o}_{T(t)}^d]^2 \tag{11}$$

where $\hat{o}_{T(t)}^d$ ignores information from the t^{th} observation, and consequently provides a measure of network performance superior to average squared error. A completely automatic method for determining network complexity appropriate for any specific application is given by choosing the number of hidden units $\hat{d}_T$ to be the smallest solution to the problem

$$\min_{d \in N_T} C_T(d) \tag{12}$$

where N_T is some appropriate choice set.

For each one-step-ahead forecast observation, the feedforward network regression is re-estimated, and the optimal network complexity is determined according to the cross-validated performance measure. Accordingly, a different model may be indicated by the cross-validated performance measure at different forecast horizons. A rolling sample approach is used, so that the same number of observations are used as the in-sample observations at every one-step-ahead prediction. The maximum number of hidden units ($N_T = 10$) and the maximum number of lags ($p = 5$) in a given feedforward regression is chosen according to the computational limitations.

4 Empirical Results

For each subsample, the forecast sample is chosen to be the last 250 days, a year of price data in daily frequency. The complete data set consists of six subsamples, a total of 1,500 observations for the forecast sample. There are two advantages to constructing the forecast sample from multiple different subsamples. The first one is to avoid spurious results as a result of data-snooping problems or sample-specific conditions. The second one is that it enables us to analyze the performance of the trading rules under different market conditions. This is particularly important in observing the performance of these rules in trendy versus sluggish market conditions in which there are no clear trends in either direction.

One-step-ahead predictions for all models are reported below. To measure the out-of-sample performance between the test regression and the benchmark model, we calculate the ratio of the respective RMSPEs. This ratio is less than 1 if the test model provides more accurate predictions. Similarly, the ratio is greater than 1 if the predictions of the test model are less accurate relative to the benchmark model. We also use a test by Mizrach (1995) to assess the statistical significance of the difference of the RMSPEs of the test regressions and

Table 2
Out-of-Sample Predictions of the Random Walk Model

1963–1967	RMSPE	0.589×10^{-2}
	σ^2	0.346×10^{-4}
1968–1971	RMSPE	0.696×10^{-2}
	σ^2	0.483×10^{-4}
1972–1975	RMSPE	0.106×10^{-1}
	σ^2	0.113×10^{-3}
1976–1979	RMSPE	0.726×10^{-2}
	σ^2	0.527×10^{-4}
1980–1983	RMSPE	0.865×10^{-2}
	σ^2	0.746×10^{-4}
1984–1986	RMSPE	0.242×10^{-1}
	σ^2	0.584×10^{-3}

Note: RMSPE refers to the the root mean square prediction error of the random walk model. σ^2 refers to the sample variance in the forecast sample.

that of the benchmark random walk model. The Mizrach (1995) test is a test for comparing predictions of univariate time series in the mean squared error.[1]

4.1 Empirical results with past returns

The RMSPEs of the random walk model and the sample variance of the forecast sample are presented in Table 2. The results with past returns are presented in Table 3. The RMSPEs of the GARCH-M (1,1), nearest neighbors, and feedforward network regression models are reported as ratios to the RMSPE of the random walk model. (The RMSPEs of the random walk model are reported in levels in Table 2.) The Mizrach (1995) test is used to calculate the statistical significance of the difference between the RMSPEs of the test and the random walk models. It is distributed as a standard normal variate under the null hypothesis so that there is no difference between the RMSPEs of the test and the benchmark[2] models.

The GARCH-M (1,1) model does not provide any forecast gain over the benchmark model. The difference between the average RMSPEs of the benchmark and the GARCH-M (1,1) models is on average less than 1%. Furthermore, none of the RMSPE test statistics are significant, indicating that the GARCH-M (1,1) model offers no forecasting improvement over the random walk model. However, the feedforward and the nearest neighbors regressions provide significant RMSPE improvements over the benchmark model. The feedforward network provides on average about 3% forecast improvement over the benchmark model. Also, the RMSPE test statistics are significant in three out of the six samples. The average RMSPE improvement of the nearest neighbors regression is also about 3%. The RMSPE statistics in this case are significant in five out of the six cases. Note that the estimation of all three models has been carried out using five lags. Using the RMSPE test, one can say that both nonparametric conditional mean estimators, the local procedure (nearest neighbors

[1]This test is very similar to the Diebold and Mariano (1995) test. For the data studied in this paper, Diebold and Mariano (1995) and Mizrach (1995) tests yield similar results. We only report the Mizrach (1995) test in the tables below.

[2]One referee also suggested that we use the Chong and Hendry (1986) encompassing test. This test is developed for forecasts generated by parametric models, and it is based on the *t*-statistic of the nesting parameter in an artificial regression that combines the two forecasts. However, in the case of two nonparametric forecasts, the asymptotic properties of this test statistic are unknown. In fact, in the context of a comparison between neural network and nearest neighbor forecasts, some preliminary results suggest that the above test over-rejects substantially, using critical values from the standard normal variate. Furthermore, in the context of our analysis, the information sets of the two models are nested and that would render any non-nested-type testing procedure inappropriate. Hence, we do not apply the above test in our empirical analysis.

Table 3
Out-of-Sample Predictions of the Models with Past Returns

		GARCH-M (1,1)	Feedforward	Nearest Neighbor
1963–1967	Ratio	0.997	0.978	0.983
	RMSPE test	0.387	1.887	1.956
1968–1971	Ratio	1.001	0.965	0.967
	RMSPE test	−0.468	2.195	2.132
1972–1975	Ratio	0.998	0.986	0.976
	RMSPE test	0.676	1.898	2.121
1976–1979	Ratio	0.997	0.981	0.972
	RMSPE test	0.854	1.876	2.121
1980–1983	Ratio	0.995	0.962	0.961
	RMSPE test	1.019	2.287	2.356
1984–1988	Ratio	0.997	0.971	0.977
	RMSPE test	1.064	2.898	2.856

Note: *Ratio* refers to the ratio of the root mean square prediction error (RMSPE) of the model in consideration to that of the random walk model. The RMSPE test is the Mizrach (1995) test statistic, and measures the statistical significance of the difference between the RMSPE of the test and the random walk models. It is distributed as a standard normal variate under the null hypothesis that there is no difference between the RMSPE of the test regression and that of the random walk model. The test statistics that are significant at the 5% level are underlined.

regression), and the global procedure (feedforward network regression) dominate their parametric counterpart, the GARCH-M (1,1) model, in forecast comparisons. Between the two, the nearest neighbors regression seems to be more successful in providing improvements over the benchmark random walk model using the RMSPE test.

4.2 Empirical results with past buy-sell signals

The predictability of the current returns with the past buy-sell signals of the moving-average rules are investigated with two different moving-average rules. These are the (1,50) and (1,200) moving-average rules. For convenience, we will call these rules A and B, respectively. The results are presented in Tables 4 and 5. All models are estimated as in the case of past returns, using five lags.

The GARCH-M(1,1) model with past buy-sell signals does not provide any forecast gain over the random walk model. The reported gains are always less than 1% forecast improvement over the random walk model. Also, none of the RMSPE test statistics are significant. The forecast gains of the feedforward and the nearest neighbors regressions are of the same order as in the case of past returns. Both models provide a forecast improvement of about 3% on average. Furthermore, five out of six for rules A and B of the RMSPE statistics were significant for the feedforward networks. All six of the RMSPE statistics were significant for the nearest neighbors regression. On the whole, both nonparametric test regressions outperform the benchmark random walk model, whereas the same cannot be said for the GARCH-M (1,1) model. Between the two nonparametric models, the nearest neighbors regression appears to be a little more successful in providing improvements over the benchmark random walk model using the RMSPE test. The comparison of the models with past returns and past buy-sell signals indicates that the latter provide more accurate forecast predictions for the current returns. Overall, the results indicate that the predictability of the current returns from the nonparametric models is statistically significant against the random walk model. Furthermore, the models with past buy-sell signals provide more consistent evidence of this predictability.

Table 4
Out-of-Sample Predictions of the Models with Rule = [1,50]

		GARCH-M (1,1)	Feedforward	Nearest Neighbor
1963–1967	Ratio	0.994	0.976	0.981
	RMSPE test	0.446	2.243	1.983
1968–1971	Ratio	0.997	0.955	0.960
	RMSPE test	0.998	2.332	1.995
1972–1975	Ratio	0.998	0.963	0.970
	RMSPE test	0.695	2.121	2.289
1976–1979	Ratio	0.994	0.975	0.971
	RMSPE test	1.176	1.798	2.112
1980–1983	Ratio	0.994	0.981	0.959
	RMSPE test	1.213	1.991	2.432
1984–1988	Ratio	0.992	0.969	0.965
	RMSPE test	1.321	2.654	2.675

Note: *Ratio* refers to the ratio of the root mean square prediction error (RMSPE) of the model in consideration to that of the random walk model. The RMSPE test is the Mizrach (1995) test statistic, and measures the statistical significance of the difference between the RMSPE of the test and the random walk models. It is distributed as a standard normal variate under the null hypothesis that there is no difference between the RMSPE of the test regression and that of the random walk model. The test statistics that are significant at the 5% level are underlined.

Table 5
Out-of-Sample Predictions of the Models with Rule = [1,200]

		GARCH-M (1,1)	Feedforward	Nearest Neighbor
1963–1967	Ratio	0.996	0.964	0.965
	RMSPE test	0.565	2.563	2.031
1968–1971	Ratio	0.995	0.965	0.959
	RMSPE test	0.676	2.307	2.012
1972–1975	Ratio	0.995	0.975	0.969
	RMSPE test	0.787	2.016	2.147
1976–1979	Ratio	1.002	0.964	0.963
	RMSPE test	−1.134	2.165	1.989
1980–1983	Ratio	0.993	0.987	0.958
	RMSPE test	1.254	1.816	2.213
1984–1988	Ratio	0.995	0.965	0.962
	RMSPE test	1.143	2.314	2.413

Note: *Ratio* refers to the ratio of the root mean square prediction error (RMSPE) of the model in consideration to that of the random walk model. The RMSPE test is the Mizrach (1995) test statistic, and measures the statistical significance of the difference between the RMSPE of the test and the random walk models. It is distributed as a standard normal variate under the null hypothesis that there is no difference between the RMSPE of the test regression and that of the random walk model. The test statistics that are significant at the 5% level are underlined.

5 Conclusions

This paper has compared the out-of-sample performances of two parametric and two nonparametric conditional mean estimators to forecast security returns with past returns and past buy-sell signals of the moving-average rules. The forecasts generated by the nonparametric models dominate the parametric ones. Among the nonparametric models, the forecasts of the local procedure (nearest neighbors regression) seems to offer more gains than the forecasts of the global procedure (feedforward regression) when measured against the benchmark random walk model using the RMSPE test. The nonparametric models that use past buy-sell signals provide more accurate RMSPEs relative to the models that use past returns. However, the reported gains may not seem to be high enough to translate into profits after transaction costs are taken into account. It would be worthwhile to carry out the above analysis as well as investigate the performance of more elaborate rules and their profitability after transaction costs and brokerage fees are taken into account. This question is left for future research.

References

Brock, W. A., J. Lakonishok, and B. LeBaron (1992). "Simple Technical Trading Rules and the Stochastic Properties of Stock Returns." *Journal of Finance* 47:1731–1764.

Chong, Y., and D. Hendry (1986). "Econometric Evaluation of Linear Macroeconomic Models." *Review of Economic Studies* LIII:671–690.

Chopra, N., J. Lakonishok, and J. R. Ritter (1992). "Performance Measurement Methodology and the Question of Whether Stocks Overreact." *Journal of Financial Economics* 31:235–268.

Colby, R. W., and T. A. Meyers (1988). *The Encyclopedia of Technical Market Indicators*. Irwin, Illinois: Business One.

Conrad, J., and G. Kaul (1988). "Time-Variation in Expected Returns." *Journal of Business* 61:409–425.

Cutler, D. M., J. M. Poterba, and L. H. Summers (1991). "Speculative Dynamics." *Review of Economic Studies* 58:529–546.

Cybenko, G. (1989). "Approximation by Superposition of a Sigmoidal Function." *Mathematics of Control, Signals and Systems* 2:303–314.

De Bondt, W. F. M., and R. H. Thaler (1985). "Does the Stock Market Overreact?" *Journal of Finance* 40:793–805.

Diebold, F. X., and R. S. Mariano (1995). "Comparing Forecasting Accuracy." *Journal of Business and Economic Statistics* 13:253–263.

Engle, R. F., D. Lilien, and R. P. Robins (1987). "Estimating Time-Varying Risk Premia in the Term Structure: The ARCH-M Model." *Econometrica* 55:391–407.

Fama, E. F. (1965). "The Behavior of Stock Market Prices." *Journal of Business* 38:34–105.

Fama, E. F. (1970). "Efficient Capital Markets: A Review of Theory and Empirical Work." *Journal of Finance* 25:383–417.

Fama, E. F. (1991). "Efficient Capital Markets: II." *Journal of Finance* 46:1575–1617.

Fama, E. F., and K. R. French (1986). "Permanent and Temporary Components of Stock Prices." *Journal of Political Economy* 98:246–274.

Fisher, L. (1966). "Some New Stock-Market Indexes." *Journal of Business* 39:191–225.

French, K. R., and R. Roll (1986). "Stock Return Variances: The Arrival of Information and the Reaction of Traders." *Journal of Financial Economics* 17:5–26.

Funahashi, K.-I. (1989). "On the Approximate Realization of Continuous Mappings by Neural Networks." *Neural Networks* 2:183–192.

Gallant, A. R., and H. White (1988). "There Exists a Neural Network that Does Not Make Avoidable Mistakes." In *Proceedings of the Second Annual IEEE Conference on Neural Networks*. New York: IEEE Press, pp. 657–664.

Gallant, A. R., and H. White (1992). "On Learning the Derivatives of an Unknown Mapping with Multilayer Feedforward Networks." *Neural Networks* 5:129–138.

Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge, UK: Cambridge University Press.

Hecht-Nielsen, R. (1989). "Theory of the Backpropagation Neural Networks." In *Proceedings of the International Joint Conference on Neural Networks.* New York: IEEE Press, pp. 593–605.

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

Hornik, K., M. Stinchcombe, and H. White (1990). "Universal Approximation of an Unknown Mapping and Its Derivatives Using Multilayer Feedforward Networks." *Neural Networks* 3:551–560.

Jegadeesh, N. (1990). "Evidence of Predictable Behavior of Security Returns." *Journal of Finance* 45:881–898.

Jensen, M. C. (1978). "Some Anomalous Evidence Regarding Market Efficiency." *Journal of Financial Economics* 6:95–101.

Kuan, C.-M., and H. White (1994). "Artificial Neural Networks: An Econometric Perspective." *Econometric Reviews* 13:1–91.

Lehmann, B. N. (1990). "Fads, Martingales and Market Efficiency." *Quarterly Journal of Economics* 105:1–28.

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.

Lo, A. W., and A. C. MacKinlay (1990). "When are Contrarian Profits Due to Stock Market Overreaction?" *Review of Financial Studies* 3:175–205.

Mizrach, B. (1995). "Forecast Comparison in L2." Working Paper 95-24, Rutgers University.

Nelson, D. B. (1991). "Conditional Heteroskedasticity in Asset Returns: A New Approach." *Econometrica* 59:347–370.

Poterba, J. M., and L. H. Summers (1988). "Mean Reversion in Stock Prices: Evidence and Implications." *Journal of Financial Economics* 22:27–59.

Robinson, P. M. (1987). "Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form." *Econometrica* 55:875–891.

Stone, C. J. (1977). "Consistent Nonparametric Regression." *Annals of Statistics* 5:595–645.

Yakowitz, S. (1987). "Nearest-Neighbor Methods for Time Series Analysis." *Journal of Time Series Analysis* 8:235–247.

Advisory Panel

Jess Benhabib, New York University

William A. Brock, University of Wisconsin-Madison

Jean-Michel Grandmont, CEPREMAP-France

Jose Scheinkman, University of Chicago

Halbert White, University of California-San Diego

Editorial Board

Bruce Mizrach (editor), Rutgers University

Michele Boldrin, University of Carlos III

Tim Bollerslev, University of Virginia

Carl Chiarella, University of Technology-Sydney

W. Davis Dechert, University of Houston

Paul De Grauwe, KU Leuven

David A. Hsieh, Duke University

Kenneth F. Kroner, BZW Barclays Global Investors

Blake LeBaron, University of Wisconsin-Madison

Stefan Mittnik, University of Kiel

Luigi Montrucchio, University of Turin

Kazuo Nishimura, Kyoto University

James Ramsey, New York University

Pietro Reichlin, Rome University

Timo Terasvirta, Stockholm School of Economics

Ruey Tsay, University of Chicago

Stanley E. Zin, Carnegie-Mellon University

Editorial Policy

The *SNDE* is formed in recognition that advances in statistics and dynamical systems theory may increase our understanding of economic and financial markets. The journal will seek both theoretical and applied papers that characterize and motivate nonlinear phenomena. Researchers will be encouraged to assist replication of empirical results by providing copies of data and programs online. Algorithms and rapid communications will also be published.

ISSN 1081-1826