Journal of International Economics 47 (1999) 91–107

Journal of
INTERNATIONAL
ECONOMICS

Linear, non-linear and essential foreign exchange rate prediction with simple technical trading rules

Ramazan Gençay*

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

Received 10 August 1996; received in revised form 21 March 1997; accepted 31 July 1997

Abstract

This paper investigates the predictability of spot foreign exchange rate returns from past buy-sell signals of the simple technical trading rules by using the nearest neighbors and the feedforward network regressions. The optimal choices for nearest neighbors, hidden units in a feedforward network and the training set are determined by the cross validation method which minimizes the mean square error. Although this method is computationally expensive the results indicate that it has the advantage of avoiding overfitting in noisy environments and indicate that simple technical rules provide significant forecast improvements for the current returns over the random walk model.

Keywords: Technical trading rules; Market timing; Non-parametric regression

JEL classification: G14; C45; C52; C53

1. Introduction

This paper studies the linear and non-linear predictability of the daily spot exchange rates from the simplest forms of technical trading rules, namely moving average rules. The moving average rule is the basis of many trend-following approaches and it is one of its simplest forms. A shorter length moving average

*Tel.: +1 519 2533000 ext. 2382; fax: +1 519 9737096; e-mail: gencay@uwindsor.ca

0022-1996/99/$ – see front matter
PII: S0022-1996(98)00017-8

rule implies a more sensitive decision rule as well as a greater number of buy and sell signals and a closer fit to the raw data. A longer length rule, on the other hand, implies a looser fit to the data, a smaller number of buy and sell signals and a greater tolerance for random movement without triggering a trade or a change in long or short positions. Taylor and Allen (1992) report the results of a questionnaire survey on technical analysis sent widely to chief foreign exchange dealers based in London, England. Taylor and Allen's findings indicate that at least 90% of the survey respondents report placing some weight on technical advice when forming their exchange rate expectations.

The single layer feedforward networks and the nearest neighbors regression are used to capture possible non-linearities in the conditional means. The first one is a global estimator, whereas the second is a local procedure that uses k-nearest neighbors to compute the estimate of the return as a weighted average. The random walk and the GARCH (1,1) models are used as linear benchmarks. The data set consists of the daily spot rates for the British pound, Deutsche mark, French franc, Japanese yen and the Swiss franc. The out-of-sample forecasts are calculated from the last one-third of each series. As a measure of performance the out-of-sample mean square prediction error (MSPE) and sign predictions are used. To assess the statistical significance of the out-of-sample predictions, the Diebold-Mariano (1995) test is calculated for all currencies. The Diebold and Mariano (1995) test is a test of the null hypothesis of no difference in the accuracy of the two competing forecasts and this test is used to evaluate the statistical significance of the MSPEs of the GARCH (1,1), the nearest neighbors and the feedforward network regression models relative to that of the random walk model.

The optimal choice of nearest neighbors, optimal number of hidden units in a feedforward network and the optimal size of the training set are determined by the cross validation method which minimizes the mean square error. As the sample moves with the forecast horizon, the cross-validated performance is recalculated to obtain the optimal number of nearest neighbors, number of hidden units and the length of the training set. Therefore, the optimal number of nearest neighbors, number of hidden units and the length of the training data set may be different corresponding to each observation in the prediction sample. The type of the cross-validation method used in this paper allows the optimal number of nearest neighbors, optimal number of hidden units and the length of the training set to be chosen dynamically. This allows non-stationarity to enter into the non-linear models in an automatic fashion.

The results of this paper indicate that simple moving average rules provide significant correct sign predictions and the Diebold and Mariano (1995) test indicates the statistical significance of these predictions when non-parametric conditional mean estimators are used to model the current returns. These results are consistent across the five currencies studied in this paper. In general, the random walk and the GARCH (1,1) models do not generate significant sign predictions and the Diebold and Mariano (1995) tests corresponding to these models are consistently insignificant. The results also indicate that cross-validation method often indicates much shorter optimal training set than the available history.

In the earlier technical trading rules literature, Dooley and Shafer (1983) and Cornell and Dietrich (1978) show that the filter type rules generate profits in excess of the buy-and-hold strategy. Later, Sweeney (1986) documents the profitability of filter rules with the Deutsche mark. Taylor (1992a), (1992b) finds similar evidence for even more extensive sets of rules and data series. Neftci (1991) designs formal algorithms to represent various forms of technical analysis in order to determine whether these rules are well defined. He concludes that trading rules provide forecast power only when the underlying process under study is non-linear. In recent papers, LeBaron (1991), (1992a) and Levich and Thomas (1993) follow the methodology of Brock et al. (1992) and use bootstrap simulations to demonstrate the statistical significance of the technical trading rules against well-known parametric null models of exchange rates.

Earlier literature has demonstrated the existence of non-linearities in the higher order moments of the foreign exchange rate series. Westerfield (1977), Boothe and Glassman (1987), Hsieh (1988) and Diebold and Nerlove (1989) all show that exchange rate changes are leptokurtic. Cumby and Obstfeld (1984), Domowitz and Hakkio (1985), Diebold (1988), McCurdy and Morgan (1988), Hsieh (1989) and Engle et al. (1990) also find conditional heteroskedasticity in the residuals of exchange rate models. The evidence of leptokurtosis or conditional heteroskedasticity, however, does not necessarily improve the forecastibility of exchange rates as these effects operate through even-ordered moments. Indeed, Meese and Rogoff (1983) show that a simple random walk model forecasts as well as most linear exchange rate models. Using non-parametric kernel regression, Diebold and Nason (1990) and Meese and Rose (1990) were not able to improve upon a simple random walk in the out-of-sample prediction of 10 major dollar spot rates in the post-1973 float period. Meese and Rose (1991) examine the five structural exchange rate models to account for potential non-linearities. Their results indicate that although non-linear effects may be important in the even moments of the exchange rate dynamics, the incorporation of non-linearities into the structural models of exchange rates do not improve our ability to understand exchange rate movements. The implication of these results constitute fairly strong evidence against the existence of asset price non-linearities that are exploitable for improved point prediction.

Contrary to earlier findings, LeBaron (1992b) finds forecast improvements of over 5% in mean squared error for the German mark. LeBaron's work relies on some results connecting volatility with conditional serial correlations of the exchange rate series. Recently, Kuan and Liu (1995) use the feedforward and recurrent network models to investigate the out-of-sample predictability of foreign exchange rates. Their results indicate that neural network models provide significantly lower out-of-sample mean squared prediction errors relative to 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 in Section 4. Conclusions follow thereafter.

2. Data description

This paper studies the daily spot rates for the British pound (BP), Deutsche mark (DM), French franc (FF), Japanese yen (JY) and the Swiss franc (SF). The data set is from the EHRA macro tape of the Federal Reserve Bank for the period of January 2, 1973 to July 7, 1992, for a total of 4894 observations. The summary statistics of the daily returns are presented in Table 1. The daily returns are calculated as the log differences of the levels. All five series exhibit slight skewness and high kurtosis which is common in high frequency financial time series data. The first 10 autocorrelations ($\rho_1, \ldots, \rho_{10}$) and the Bartlett standard errors from these series are also reported in Table 1. All series show evidence of autocorrelation. 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

Table 1
Summary statistics for the daily exchange rates: log first difference January 2, 1973—July 7, 1992

Description	FF	DM	JY	SF	BP
Sample size	4893				
Mean*100	0.000291	0.015520	0.018081	0.021019	−0.004055
Std.*100	0.662688	0.679150	0.622425	0.771470	0.634438
Skewness	−0.107067	0.061274	0.945698	−0.062692	−0.082874
Kurtosis	10.498755	8.384598	19.753822	6.724369	7.085346
Max	0.060490	0.061985	0.095073	0.044083	0.045885
Min	−0.058744	−0.058678	−0.062566	−0.058269	−0.038427
ρ_1	0.035102	0.041546	0.049731	0.038438	0.066912
ρ_2	−0.011794	−0.016066	0.012342	−0.002566	−0.008199
ρ_3	0.019718	0.024828	0.036739	0.008828	−0.010606
ρ_4	0.000272	−0.010582	−0.004319	−0.012799	−0.006968
ρ_5	0.022906	0.029296	0.030632	0.015076	0.046339
ρ_6	−0.009401	0.015755	0.008076	0.017570	−0.010261
ρ_7	0.022866	0.004933	0.014625	−0.004060	−0.011901
ρ_8	0.028023	0.037632	0.010479	0.025496	0.011131
ρ_9	0.043241	0.039234	0.036994	0.033734	0.035486
ρ_{10}	0.010825	0.013774	0.038678	0.016768	−0.022401
Bartlett std. errors	0.014295				
LBP	27.6	33.0	38.3	20.3	44.0
$\chi^2_{0.05}(10)$	18.307				

FF, DM, JY, SF and BP refer to French franc, Deutsche mark, Japanese yen, Swiss franc and British pound, respectively. $\rho_1, \ldots, \rho_{10}$ 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.

hypothesis of identical and independent observations. All five series reject the null hypothesis of identical and independent observations.

3. Estimation techniques

Let p_t, $t=1,2,\ldots,T$ be the daily spot exchange rate 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

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

The buy and sell signals are calculated by

$$s_t^{n1,n2}=m_t^{n1}-m_t^{n2} \tag{2}$$

where $n1$ and $n2$ are the short and the long moving averages, respectively. The rules used in this paper are $(n1,n2)=[(1,50),(1,200)]$ where $n1$ and $n2$ are in days.

3.1. Parametric test model

The GARCH (1,1) specification (Engle (1982) and Bollerslev (1986)) is used as the parametric test model

$$r_t=\alpha+\sum_{i=1}^{p}\beta_i s_{t-i}^{n1,n2}+\epsilon_t \tag{3}$$

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 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.

There are numerous non-parametric regression techniques available such as flexible fourier forms, nearest neighbors regression, 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 and nearest neighbors regression models are used. These two non-parametric regression models are described below.

3.2. 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 non-parametric 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 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 explained in the following way. Take the time series $\{x_t\}_{t=1}^{T}$ and convert it into a series 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 the ω_{ti}s are the k nearest neighbor weights. 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 uniform weights are used to weigh the contribution of the k nearest neighbors in the overall estimate of the regression function. In the context of the present application, the regression function is written as

$$E(r_t|s_{t-1}^{n1,n2},s_{t-2}^{n1,n2},\ldots,s_{t-p}^{n1,n2})=\sum_{i=1}^{k}\omega_{ti}r_i \tag{4}$$

where $p=9$ and $k=\{1,2,\ldots,100\}$

The optimal choice of nearest neighbors was determined by the cross validation method which minimizes the mean square error. As the sample moves with the forecast horizon, the cross-validated performance is recalculated to obtain the optimal number of nearest neighbors. Therefore, the optimal number of nearest neighbors may be different corresponding to each observation in the prediction sample. In Diebold and Nason (1990), the predictions are based on a given number of nearest neighbors over the entire forecast sample and the number of nearest neighbors is not allowed to vary as the sample[1] moves with the progression of the

[1]The cross-validation procedure of this paper is also studied with the entire in-sample history to forecast the last one-third of the spot exchange returns, using the lagged returns as inputs. Using the entire history degrades the results of this paper and the sign predictions go down to the 50% level with the non-parametric methods.

forecast iterations. In Kuan and Liu (1995), an information theoretic criteria (Prediction Stochastic Complexity) is used to determine the number of hidden units in a network. Once the optimal network structure is determined, it is fixed throughout the entire forecast sample.

For the implementation of the cross-validation, the 250 most recent observations (~1 year of trading data) in the in-sample data (training set) are used first to calculate the optimal number of nearest neighbors corresponding to the smallest cross-validated mean square error (MSE) among all possible nearest neighbors. Afterwards, one observation from the further past is added to the set of 250 observations which brings up the total number of observations to 251. The cross-validation takes place in order to calculate the optimal number of nearest neighbors corresponding to the smallest cross-validated mean square error among all possible nearest neighbors.

This process of adding observations further from the past is carried out until all in-sample observations are utilized. This process gives us a sequence of optimal nearest neighbors corresponding to each in-sample data length. From this sequence, the number of nearest neighbors corresponding to the smallest MSE is determined. This choice of the number of nearest neighbors also corresponds to a certain in-sample data length. These choices for the number of nearest neighbors and the in-sample data length are used to predict the first available forecast observation[2] and to calculate the corresponding forecast error.

In the next step, the sample is rolled one observation forward and the second observation in the forecast sample is studied in the same manner as described for the first forecast observation above. This procedure is repeated until all observations are covered in the forecast sample. In the final step, sign predictions and the Diebold and Mariano (1995) tests are calculated based on the mean square prediction errors (MSPEs) of the test and the random walk models.

My data methodology is clearly an expensive one. It however has the following advantages.

- The number of nearest neighbors and the number of observations needed for the in-sample estimation are determined optimally which prevents overfitting in noisy environments. In other words, this procedure may utilize only a certain

[2]For instance, to forecast $t+1$, the system starts by looking at the previous 250 points. For each k ($k=1,2,\ldots,100$), leave-one-out cross validation is done with nine lags as inputs. The corresponding k and the *MSE* are recorded for k ($k=1,2,\ldots,100$). The k which corresponds to the smallest MSE is chosen for the training size of 250 observations. The history is then extended by one into the past, and the procedure is repeated. Similar to the above, the k which corresponds to the smallest MSE is selected for the set of 251 training observations. This procedure is repeated until the complete training set is reached. This gives a sequence of k's which correspond to the smallest MSE for each training data length. Finally, the k and the corresponding training data length are selected among the sequence of k's of different training data lengths.

number of in-sample observations rather than the entire in-sample if a certain subset of the in-sample observations provide smaller MSE relative to the MSE of the entire in-sample set. This allows non-stationarity to enter into the models in an automatic fashion.

- It is a fair procedure as it only relies on in-sample performance. The parameter estimates used for forecasting purposes are obtained from the in-sample estimation and the observations in the forecast sample (out-of-sample observations) are not utilized at any stage of the model specification or estimation. Finally, this model selection methodology is completely data driven.

3.3. Feedforward networks

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

$$r_t = \alpha_0 + \sum_{j=1}^{d} \beta_j G\left(\alpha_j + \sum_{i=1}^{p} \gamma_{ij} s_{t-i}^{n1,n2}\right) + \epsilon_t \quad \epsilon_t \sim ID(0,\sigma_t^2) \tag{5}$$

where G is the known activation function which 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; Funanhashi, 1989; Hecht-Nielsen, 1989; Hornik et al., 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 (3), (4) and (5) the random walk model

$$r_t = \alpha + \epsilon_t \quad \epsilon_t \sim ID(0,\sigma^2) \tag{6}$$

is used as the benchmark model. The out-of-sample forecast performance of Eq. (3), Eq. (4) and Eq. (5) are measured by the ratio of their MSPEs to that of the random walk model in Eq. (6). 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 MSPE of the GARCH (1,1) model with lagged returns

$$r_t = \alpha + \sum_{i=1}^{p} \beta_i r_{t-i} + \epsilon_t \quad \epsilon_t \sim N(0,h_t) \quad h_t = \delta_0 + \delta_1 h_{t-1} + \delta_2 \epsilon_{t-1}^2 \tag{7}$$

is compared to that of the benchmark model in Eq. (6). The out-of-sample forecast performance of the single layer feedforward network model with lagged returns

$$r_t = \alpha_0 + \sum_{j=1}^{d} \beta_j G\left(\alpha_j + \sum_{i=1}^{p} \gamma_{ij} r_{t-i}\right) + \epsilon_t \quad \epsilon_t \sim ID(0,\sigma_t^2) \tag{8}$$

is also compared to that of the benchmark model in Eq. (6). Feedforward network regression models require a choice for the number of hidden units in a network which is determined by a cross-validation method. The cross-validation method used for the feedforward network models is the same as the one described for nearest neighbors regression except that the number of hidden units is the choice variable for the feedforward network regression. The set of number of hidden units is chosen to be $\{1,2,\ldots,15\}$ and the number of lags is set to $p=9$ to capture the potential persistence in the series.

In the implementation of the cross-validation for the parametric models, the 250 most recent training observations are used first to calculate the cross-validated mean square error (MSE). Afterwards, one observation from the further past is added to the set of 250 observations which brings up the total number of observations to 251. The cross-validation takes place in order to calculate the cross-validated mean square error. This process of adding observations further from the past is carried out until all in-sample observations are utilized. This process gives us a sequence of mean squared errors and the corresponding training data length. From this sequence, the data length which corresponds to the smallest mean squared error is chosen to predict the first available forecast observation and to calculate the corresponding forecast error. In the next step, the sample is rolled one observation forward and the second observation in the forecast sample is studied in the same manner as described for the first forecast observation above. This procedure is repeated until all observations are covered in the forecast sample. In the final step, sign predictions and the Diebold and Mariano (1995) tests are calculated based on the mean square prediction errors (MSPEs) of the test and the random walk models.

4. Empirical results

For each exchange rate the out-of-sample predictive performance of the parametric and the non-parametric conditional mean estimators is examined. The data set consists of daily spot rates for the British pound (BP), Deutsche mark (DM), French franc (FF), Japanese yen (JY) and Swiss franc (SF). The out-of-sample forecasts are calculated from the last one-third of the data set. In total, there are 4684 observations for estimation so that the last 1561 observations are kept for the out-of-sample predictions. As a measure of performance the out-of-sample mean square prediction error (MSPE) and sign predictions are used. To assess the statistical significance of the out-of-sample predictions, the Diebold-Mariano (1995) test is calculated for all currencies. The Diebold and Mariano (1995) test is a test of the null hypothesis of no difference in the accuracy of the two competing forecasts and this test is used to evaluate the statistical significance of the MSPEs of the GARCH (1,1), nearest neighbors and feedforward network regression models relative to that of the random walk model. In addition to the Diebold-Mariano (1995) test, the percentage correct sign predictions of the out-of-sample forecasts are reported.

Out-of-sample forecasts are completely ex ante, using only information actually available. The out-of-sample forecasts with the buy-sell signals as the conditioning set are computed recursively by estimating $E(r_t|s_{t-1}^{n1,n2},\ldots,s_{t-p}^{n1,n2})$ and then $E(r_{t+1}|s_t^{n1,n2},\ldots,s_{t-p+1}^{n1,n2})$ and so forth, in real time. The out-of-sample forecasts with past returns as the conditioning set are computed in a similar fashion.[3]

The results with past returns are presented in Table 2. The mean square prediction errors (MSPEs) of the random walk model in the first panel of Table 2 are obtained from the parameters estimated from the entire training period. MSPE ratio[4] refers to the ratio of the MSPE of the corresponding model to that of the random walk model. The MSPE ratio is less than one if the model under consideration provides smaller MSPE relative to the MSPE of the random walk model. In the MSPE ratio calculations the MSPE of the random walk model is calculated from the optimal training data set indicated by the cross-validation procedure for the GARCH (1,1) and the non-parametric models. For instance, if the cross-validation indicates to use 253 most recent observations from the training set for the nearest neighbors regression, the same training set is used to calculate the forecast of the random walk model. This ensures that the forecast comparisons are fair between the random walk model and the model in comparison as they both use the same training set for their forecasts.[5]

D&M refers to the Diebold and Mariano (1995) test for a mean loss differential. This test statistic is distributed standard normal in large samples. All D&M test

[3]To account for the days-of-the-week effects, return series are also filtered by

$$r_t = \alpha + \beta_1 D_{MO,t} + \beta_2 D_{TU,t} + \beta_3 D_{WE,t} + \beta_4 D_{TH,t} + \epsilon_t$$

where $D_{MO,t}$, $D_{TU,t}$, $D_{WE,t}$ and $D_{TH,t}$ are the dummy variables for Monday, Tuesday, Wednesday and Thursday for observation t. The residuals from this regression are used as the filtered returns. The results with the filtered returns are similar to that of the unfiltered returns.

[4]In the paper, the GARCH (1,1) and the two non-parametric models are compared to the random walk model. This serves as the first benchmark. The MSPE ratios between GARCH (1,1) and the non-parametric models can also be compared which gives the forecast performance between GARCH (1,1) and the non-parametric models. This can be considered as the second linear benchmark of the paper.

[5]When the random walk forecasts from the full training set are compared to the GARCH (1,1) and the non-parametric models with cross-validation, the models in comparison do slightly better relative to the random walk model.

Table 2
Out-of-sample prediction results with past returns

Models	Statistics	BP	DM	FF	JY	SF
Random walk						
	MSPE	0.455	0.493	0.452	0.434	0.591
	Sign	0.47	0.51	0.48	0.50	0.50
GARCH (1,1)						
	MSPE ratio	0.996	0.997	0.995	0.991	0.990
	D&M	0.156	0.167	0.163	0.149	0.140
	Sign	0.50	0.51	0.51	0.51	0.50
Feedforward networks						
	MSPE ratio	0.917	0.916	0.923	0.909	0.939
	D&M	0.053	0.052	0.057	<u>0.047</u>	0.052
	Sign	0.55	0.56	0.54	0.56	0.55
	Hidden units	7	8	8	7	9
Nearest neighbors						
	MSPE ratio	0.903	0.907	0.911	0.895	0.898
	D&M	<u>0.037</u>	<u>0.023</u>	<u>0.032</u>	<u>0.038</u>	<u>0.031</u>
	Sign	0.58	0.58	0.59	0.59	0.55
	Nearest neighbors	16	19	18	13	19

MSPE is the mean square prediction error of the random walk model and it is reported in levels ($\times 10^4$). MSPE ratio is the ratio of the MSPE of the corresponding model to that of the random walk model. D&M refers to the Diebold and Mariano (1995) test for a mean loss differential. This test statistic is distributed standard normal in large samples. All D&M test statistics are calculated from the loss differential of the mean square prediction errors of the corresponding model to that of the random walk model. In the table above, the *P*-values of the D&M statistics are reported and underlined if less than 5%. Sign refers to the percentage of the correct signs in the out-of-sample period. For the nearest neighbors and feedforward regressions, the average number of nearest neighbors and the number of hidden units from the in-sample estimation are reported in the last rows of the corresponding panels for each method.

statistics are calculated from the loss differential of the mean square prediction errors of the corresponding model to that of the random walk model. The *P*-values of the D&M test are reported in the tables and underlined if it is less than 5%. Sign refers to the percentage of the correct sign predictions in the out-of-sample period.

The results of the random walk model indicate that sign predictions average around the 49% level. For the GARCH-M (1,1) model, the MSPE ratios are close to one and sign predictions are 50, 51, 51, 51 and 50% for the BP, DM, FF, JY and SF, respectively. The MSPE ratios indicate that the feedforward network model provides an average of 7.9% forecast improvement over the random walk model across all currencies. The feedforward network model also provides more accurate sign predictions relative to the random walk model although the statistical significance of the D&M tests are at the margin at the 5% level. The nearest neighbors model provides significant forecast gains over both the parametric and the feedforward network models. The average forecast gain for the nearest neighbors regression is 9.7% across all five currencies. Sign predictions of the nearest neighbors regression are 58, 58, 59, 59 and 55% for the BP, DM, FF, JY and SF, respectively.

For the non-parametric conditional mean estimators, the local procedure (nearest neighbors regression) dominate the global procedure (feedforward network regression) in forecast comparisons. The results of the nearest neighbor regression shows that the average *k* that minimized the mean square error is 17 across the five currencies. For the feedforward network regression, the average number of hidden units is 8 for the five currencies.

In Table 3 and Table 4, the predictability of the current returns with the past buy-sell signals of the moving average rules are investigated with the (1,50) and (1,200) rules.[6] Relative to the results in Table 2, GARCH (1,1) model provides a slight improvement for the BP, DM and FF series. For the JY and SF, the performance of the GARCH (1,1) model is worse than the ones in Table 2 when

Table 3
Out-of-sample prediction results with rule MA = [1,50]

Models	Statistics	BP	DM	FF	JY	SF
GARCH (1,1)						
	MSPE ratio	0.993	0.992	0.994	0.995	0.994
	D&M	0.145	0.143	0.156	0.152	0.138
	Sign	0.51	0.50	0.49	0.49	0.51
Feedforward networks						
	MSPE ratio	0.893	0.890	0.895	0.880	0.887
	D&M	0.007	0.006	0.011	0.007	0.008
	Sign	0.59	0.57	0.58	0.57	0.59
	Hidden units	9	6	8	7	8
Nearest neighbors						
	MSPE ratio	0.875	0.881	0.856	0.861	0.872
	D&M	0.004	0.005	0.005	0.008	0.007
	Sign	0.62	0.63	0.63	0.62	0.61
	Nearest neighbors	19	17	20	16	17

MSPE ratio is the ratio of the MSPE of the corresponding model to that of the random walk model. D&M refers to the Diebold and Mariano (1995) test for a mean loss differential. This test statistic is distributed standard normal in large samples. All D&M test statistics are calculated from the loss differential of the mean square prediction errors of the corresponding model to that of the random walk model. In the table above, the *P*-values of the D&M statistics are reported and underlined if less than 5%. Sign refers to the percentage of the correct signs in the out-of-sample period. For the nearest neighbors and feedforward regressions, the average number of nearest neighbors and the number of hidden units from the in-sample estimation are reported in last rows of the corresponding panels for each method.

[6] I also studied the linear and non-parametric models when the technical trading indicators are in the form of 1 and 0's for the buy and sell signals. The results indicate that both the linear and the non-parametric models yield forecasts predictions which are within 1% of the forecast predictions obtained with the continuous buy-sell signals that are presented in Table 3 and Table 4.

Table 4
Out-of-sample prediction results with rule MA = [1,200]

Models	Statistics	BP	DM	FF	JY	SF
GARCH (1,1)						
	MSPE ratio	0.987	0.991	0.992	0.993	0.991
	D&M	0.134	0.146	0.154	0.149	0.142
	Sign	0.52	0.51	0.50	0.49	0.51
Feedforward networks						
	MSPE ratio	0.897	0.891	0.893	0.883	0.886
	D&M	0.006	0.006	0.009	0.008	0.007
	Sign	0.58	0.57	0.57	0.57	0.58
	Hidden units	8	7	8	9	10
Nearest neighbors						
	MSPE ratio	0.884	0.883	0.861	0.867	0.879
	D&M	0.005	0.007	0.006	0.007	0.006
	Sign	0.61	0.60	0.61	0.63	0.59
	Nearest neighbors	17	18	19	18	16

MSPE ratio is the ratio of the MSPE of the corresponding model to that of the random walk model. D&M refers to the Diebold and Mariano (1995) test for a mean loss differential. This test statistic is distributed standard normal in large samples. All D&M test statistics are calculated from the loss differential of the mean square prediction errors of the corresponding model to that of the random walk model. In the table above, the *P*-values of the D&M statistics are reported and underlined if less than 5%. Sign refers to the percentage of the correct signs in the out-of-sample period. For the nearest neighbors and feedforward regressions, the average number of nearest neighbors and the number of hidden units from the in-sample estimation are reported in last rows of the corresponding panels for each method.

the MSPE ratios are compared. The sign predictions of the GARCH (1,1) model in Table 3 and Table 4 are comparable to the GARCH (1,1) sign predictions in Table 2. The performance of the feedforward and nearest neighbors models both indicate significantly lower MSPEs relative to the random walk model and significantly higher sign predictions. With rule (1,50) for BP, DM, FF, JY and SF, the feedforward network provides 10.7, 11.0, 10.5, 12.0 and 11.3% forecast improvement over the random walk model when the MSPEs of both models are compared. For the nearest neighbors regression the MSPEs are 12.5, 11.9, 14.4, 13.9 and 12.8% smaller than the random walk model for all five currencies. The feedforward network provides an average of 58% correct sign predictions for the five currencies. The average sign prediction of the nearest neighbors model is 62% for the five currencies. The comparison of the nearest neighbors and the feedforward network models indicate that the nearest neighbors models do slightly better than the feedforward network models. This suggests that feedforward networks may be suffering from a degree of oversmoothing relative to the nearest neighbors estimates. Overall, both types of non-parametric forecasts outperform the GARCH (1,1) and the random walk model forecasts. The forecasts are more accurate when the past buy-sell signals are used relative to the past returns as inputs. The comparison of the (1,50) rule with (1,200) also indicates that the (1,50) rule provides more accurate forecasts over the (1,200) rule. This may be due to the fact that the (1,200) oversmooths the data. The results of the nearest neighbor regression with the past buy-sell signals show that the average k that minimized the mean square error is 18 for rules (1,50) and (1,200) across five currencies. For the feedforward network regression, the average number of hidden units is 8 across all five currencies.

In addition to the one-step ahead predictions, 5- and 10-step ahead predictions are also studied with past returns and past buy-sell signals of the moving average rules. At the 5- and the 10-step prediction levels, there is no evidence of predictability for all currencies when the current returns are modelled by past returns or past buy-sell signals. This suggests that the predictive power of the buy-sell signals are limited to the immediate future (which is the next day with the daily data used in this paper) and these rules may not have long-term predictive ability.

The cross-validation methodology in this paper starts from the most recent 250 observations in the training set and searches the entire training set to obtain the optimal training data length within the context of an econometric model. Therefore, a training data size as small as 250 observations to the entire training set are studied. The cross-validation results of this paper indicate that optimal training data length is much less than the entire training size. The average optimal training data length for all currencies is an average of one-third of the total training set of 1561 observations. This is consistent across the GARCH (1,1) and the non-parametric models. This means that larger training sets may lead to overfitting and therefore poor out-of-sample generalizations.

5. Conclusions

This paper has compared the out-of-sample performances of two parametric and two non-parametric conditional mean estimators to forecast spot exchange rate returns with past returns and past buy-sell signals of the moving average rules. The forecasts generated by the non-parametric models dominate the parametric ones. Among the non-parametric models, the forecasts of the local procedure (nearest neighbors regression) dominate the global procedure (feedforward network regression). This suggests that feedforward networks may be suffering from a degree of oversmoothing relative to the nearest neighbors estimates.

The results indicate that simple moving average rules provide significant correct sign predictions and the Diebold and Mariano (1995) test indicates the statistical significance of these predictions when non-parametric conditional mean estimators are used to model the current returns. These results are consistent across the five currencies studied in this paper. In general, the random walk and the GARCH (1,1) models do not generate significant sign predictions and Diebold and Mariano (1995) tests corresponding to these models are consistently insignificant.

This paper has used cross-validation as a model and data selection procedure. Although this method is computationally expensive, it has the advantage that the model complexity and the number of observations needed for the in-sample estimation are determined optimally which prevents overfitting in noisy environments. In other words, this procedure may utilize only a certain number of in-sample observations rather than the entire in-sample if a certain subset of the in-sample observations provides smaller mean square error relative to the mean square error of the entire in-sample set. It is a fair procedure as it only relies on in-sample performance. In other words, observations in the forecast sample (out-of-sample observations) are not utilized at any stage of the model specification.

Acknowledgements

I thank the editor and two anonymous referees for their suggestions. I also thank the Natural Sciences and Engineering Research Council of Canada and the Social Sciences and Humanities Research Council of Canada for financial support.

References

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Boothe, P., Glassman, D., 1987. The statistical distribution of exchange rates. Journal of International Economics 22, 297–319.

Brock, W.A., Lakonishok, J., LeBaron, B., 1992. Simple technical trading rules and the stochastic properties of stock returns. Journal of Finance 47, 1731–1764.

Cornell, W.B., Dietrich, J.K., 1978. The efficiency of the market for foreign exchange under floating exchange rates. Review of Economics and Statistics 60, 111–120.

Cumby, R., Obstfeld, M., 1984. International interest-rate and price-level linkages under flexible exchange rate: a review of recent evidence. In: Bilson, J.F.O., Marston, R. (Eds.), Exchange Rate Theory and Practice. University of Chicago Press, Chicago, pp. 121–151.

Cybenko, G., 1989. Approximation by superposition of a sigmoidal function. Mathematics of Control, Signals and Systems 2, 303–314.

Diebold, F.X., 1988. Empirical Modelling of Exchange Rate Dynamics. Springer-Verlag, New York.

Diebold, F.X., Nason, J.A., 1990. Nonparametric exchange rate prediction? Journal of International Economics 28, 315–332.

Diebold, F.X., Nerlove, M., 1989. The dynamics of exchange rate volatility: a multivariate latent-factor ARCH model. Journal of Applied Econometrics 4, 1–22.

Diebold, F.X., Mariano, 1995. Comparing forecasting accuracy, Journal of Business and Economic Statistics 13, 253–263.

Domowitz, I., Hakkio, C.S., 1985. Conditional variance and the risk premium in the foreign exchange market. Journal of International Economics 19, 47–66.

Dooley, M.P., Shafer, J., 1983. Analysis of short-run exchange rate behavior: March 1973 to November 1981. In: Bigman, D., Taya, T. (Eds.), Exchange Rate and Trade Instability: Causes, Consequences and Remedies. Ballinger, Cambridge, pp. 43–72.

Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation . Econometrica 50, 987–1008.
Engle, R.F., Ito, T., Lin, W., 1990. Meteor showers or heat waves? Heteroskedastic daily volatility in the foreign exchange market. Econometrica 58, 525–542.
Funanhashi, K.-I., 1989. On the approximate realization of continuous mappings by neural networks. Neural Networks 2, 183–192.
Gallant, A.R., White, H., 1988. There exists a neural network that does not make avoidable mistakes. Proceedings of the second annual IEEE conference on neural networks, San Diego, CA, vol. 1. IEEE Press, New York, pp. 657–664.
Gallant, A.R., White, H., 1992. On learning the derivatives of an unknown mapping with multilayer feedforward networks. Neural Networks 5, 129–138.
Härdle, W., 1990. Applied non-parametric regression. Cambridge University Press, New York.
Hecht-Nielsen, R., 1989. Theory of the backpropagation neural networks. Proceedings of the international joint conference on neural networks, Washington DC, vol. 1. IEEE Press, New York, pp. 593–605.
Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Networks 2, 359–366.
Hornik, K., Stinchcombe, M., White, H., 1990. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. Neural Networks 3, 551–560.
Hsieh, D.A., 1988. The statistical properties of daily foreign exchange rates: 1974–1983. Journal of International Economics 24, 129–145.
Hsieh, D.A., 1989. Testing for nonlinear dependence in foreign exchange rates: 1974–1983. Journal of Business 62, 339–368.
Kuan, C.-M., Liu, T., 1995. Forecasting exchange rates using feedforward and recurrent neural networks. Journal of Applied Econometrics 10, 347–364.
Kuan, C.-M., White, H., 1994. Artificial neural networks: an econometric perspective. Econometric Reviews 13, 1–91.
LeBaron, B., 1991. Technical trading rules and regime shifts in foreign exchange. Social Systems Research Institute, University of Wisconsin-Madison.
LeBaron, B., 1992a. Do moving average trading rule results imply nonlinearities in foreign exchange markets? Social Systems Research Institute, University of Wisconsin-Madison.
LeBaron, B., 1992b. Forecast improvements using a volatility index. Journal of Applied Econometrics 7, 137–150.
Levich, R.M., Thomas, L.R., 1993. The significance of technical trading-rule profits in the foreign exchange market: a bootstrap approach. Journal of International Money and Finance 12, 451–474.
McCurdy, T.H., Morgan, I.G., 1988. Testing the martingale hypothesis in Deutsche Mark futures with models specifying the form of heteroskedasticity. Journal of Applied Econometrics 3, 187–202.
Meese, R.A., Rogoff, K., 1983. Empirical exchange rate models of the seventies: do they fit out sample? Journal of International Economics 14, 3–24.
Meese, R.A., Rose, A.K., 1990. Non-linear, non-parametric, non-essential exchange rate estimation. American Economic Review 80, 192–196.
Meese, R.A., Rose, A.K., 1991. An empirical assessment of non-linearities in models of exchange rate determination. Review of Economic Studies 58, 603–619.
Neftci, S.N., 1991. Naive trading rules in financial markets and Wiener-Kolmogorov prediction theory: a study of 'Technical Analysis'. Journal of Business 64, 549–571.
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.
Sweeney, R.J., 1986. Beating the foreign exchange market. Journal of Finance 41, 163–182.
Taylor, M.P., Allen, H., 1992. The use of technical analysis in the foreign exchange market. Journal of International Money and Finance 11, 304–314.

Taylor, S.J., 1992. Rewards available to currency futures speculators: compensation for risk or evidence of inefficient pricing? Economic Record 68, 105–116.

Taylor, S.J., 1992b. Trading futures using the channel rule: a study of the predictive power of technical analysis with currency examples. The Management School, Lancaster University, UK.

Westerfield, J.M., 1977. An examination of foreign exchange risk under fixed and floating rate regimes. Journal of International Economics 7, 181–200.

Yakowitz, S., 1987. Nearest-neighbor methods for time series analysis. Journal of Time Series Analysis 8, 235–247.