Journal of Forecasting
J. Forecast. **21**, 317–354 (2002)
Published online 11 June 2002 in Wiley InterScience (www.interscience.wiley.com). **DOI**: 10.1002/for.833

Forecasting and Trading Currency Volatility: An Application of Recurrent Neural Regression and Model Combination

CHRISTIAN L. DUNIS* AND XUEHUAN HUANG
Liverpool Business School and CIBEF, UK

ABSTRACT

In this paper, we examine the use of non-parametric Neural Network Regression (NNR) and Recurrent Neural Network (RNN) regression models for *forecasting* and *trading* currency volatility, with an application to the GBP/USD and USD/JPY exchange rates. Both the results of the NNR and RNN models are *benchmarked* against the simpler GARCH alternative and implied volatility. Two simple model combinations are also analysed.

The intuitively appealing idea of developing a nonlinear nonparametric approach to forecast FX volatility, identify mispriced options and subsequently develop a trading strategy based upon this process is implemented for the first time on a comprehensive basis. Using daily data from December 1993 through April 1999, we develop alternative FX volatility forecasting models. These models are then tested *out-of-sample* over the period April 1999–May 2000, not only in terms of *forecasting accuracy*, but also in terms of *trading efficiency*: in order to do so, we apply a realistic volatility trading strategy using FX option straddles once mispriced options have been identified.

Allowing for transaction costs, most trading strategies retained produce positive returns. RNN models appear as the best single modelling approach yet, somewhat surprisingly, model combination which has the best overall performance in terms of forecasting accuracy, fails to improve the RNN-based volatility trading results.

Another conclusion from our results is that, for the period and currencies considered, the currency option market was inefficient and/or the pricing formulae applied by market participants were inadequate.

KEY WORDS forecasting accuracy; model combination; recurrent neural networks; trading efficiency; volatility modelling

INTRODUCTION

Exchange rate volatility has been a constant feature of the International Monetary System ever since the breakdown of the Bretton Woods system of fixed parities in 1971–3. Not surprisingly, in the

* Correspondence to: Christian L. Dunis, 32 Montholme Road, London, SW11 6HY, UK. E-mail: cdunis@totalise.co.uk

wake of the growing use of derivatives in other financial markets, and following the extension of the seminal work of Black and Scholes (1973) to foreign exchange by Garman and Kohlagen (1983), currency options have become an ever more popular way to hedge foreign exchange exposures and/or speculate in the currency markets.

In the context of this wide use of currency options by market participants, having the best volatility prediction has become ever more crucial. True, the only unknown variable in the Garman–Kohlagen pricing formula is precisely the future foreign exchange rate volatility during the life of the option. With an 'accurate' volatility estimate and knowing the other variables (strike level, current level of the exchange rate, interest rates on both currencies and maturity of the option), it is possible to derive the theoretical arbitrage-free price of the option. Just because there will never be such thing as a unanimous agreement on the future volatility estimate, market participants with a better view/forecast of the evolution of volatility will have an edge over their competitors.

In a rational market, the equilibrium price of an option will be affected by changes in volatility. The higher the volatility perceived by market participants, the higher the option's price. Higher volatility implies a greater possible dispersion of the foreign exchange rate at expiry: all other things being equal, the option holder has logically an asset with a greater chance of a more profitable exercise. In practice, those investors/market participants who can reliably predict volatility should be able to control better the financial risks associated with their option positions and, at the same time, profit from their superior forecasting ability.

There is a wealth of articles on predicting volatility in the foreign exchange market: for instance, Baillie and Bollerslev (1990) used ARIMA and GARCH models to describe the volatility on hourly data, West and Cho (1995) analysed the predictive ability of GARCH, AR and non-parametric models on weekly data, Jorion (1995) examined the predictive power of implied standard deviation as a volatility forecasting tool with daily data, Dunis *et al.* (2001b) measure, using daily data, both the 1-month and 3-month forecasting ability of 13 different volatility models including AR, GARCH, stochastic variance and model combinations with and without the adding of implied volatility as an extra explanatory variable.

Nevertheless, with the exception of Engle *et al.* (1993), Dunis and Gavridis (1997) and, more recently, Laws and Gidman (2000), these papers evaluate the out-of-sample forecasting performance of their models using traditional statistical accuracy criteria, such as root mean squared error, mean absolute error, mean absolute percentage error, Theil-*U* statistic and correct directional change prediction. Investors and market participants however have trading performance as their ultimate goal and will select a forecasting model based on financial criteria rather than on some statistical criterion such as root mean squared error minimization. Yet, as mentioned above, seldom has recently published research applied any financial utility criterion in assessing the out-of-sample performance of volatility models.

Over the past few years, Neural Network Regression (NNR) has been widely advocated as a new alternative modelling technology to more traditional econometric and statistical approaches, claiming increasing success in the fields of economic and financial forecasting. This has resulted in many publications comparing neural networks and traditional forecasting approaches. In the case of foreign exchange markets, it is worth pointing out that most of the published research has focused on exchange rate forecasting rather than on currency volatility forecasts. However, financial criteria, such as Sharpe ratio, profitability, return on equity, maximum drawdown, etc., have been widely used to measure and quantify the out-of-sample forecasting performance. Dunis (1996) investigated the application of NNR to intraday foreign exchange forecasting and his results were evaluated by means of a trading strategy. Kuan and Liu (1995) proposed two-step Recurrent Neural Network

(RNN) models to forecast exchange rates and their results were evaluated using traditional statistical accuracy criteria. Tenti (1996) applied RNNs to predict the USD/DEM exchange rate, devising a trading strategy to assess his results, while Franses and Van Homelen (1998) use NNR models to predict four daily exchange rate returns relative to the Dutch guilder using directional accuracy to assess out-of-sample forecasting accuracy. Overall, it seems, however, that neural network research applied to exchange rates has been so far seldom devoted to FX volatility forecasting.

Accordingly, the rationale for this paper is to investigate the predictive power of alternative non-parametric forecasting models of foreign exchange volatility, from both a *statistical* and an *economic* point of view. We examine the use of Neural Network Regression (NNR) and Recurrent Neural Network (RNN) regression models for *forecasting* and *trading* currency volatility, with an application to the GBP/USD and USD/JPY exchange rates. The results of the NNR and RNN models are *benchmarked* against the simpler GARCH(1,1) alternative, implied volatility and model combinations: in terms of model combination, a simple average combination and the Granger/Ramanathan (1984) optimal weighting regression-based approach are employed and their results investigated.

Using daily data from December 1993 through April 1999, we develop alternative FX volatility forecasting models. These models are then tested out-of-sample over the period April 1999–May 2000, not only in terms of forecasting accuracy but also in terms of *trading efficiency*. In order to do so, we apply a realistic volatility trading strategy using FX option straddles once mispriced options have been identified.

Allowing for transaction costs, most trading strategies retained produce positive returns. RNN models appear as the best single modelling approach but model combinations, despite their superior performance in terms of forecasting accuracy, fail to produce superior trading strategies.

Another conclusion from our results is that, for the period and currencies considered, the currency option market was inefficient and/or the pricing formulae applied by market participants were inadequate.

Overall, we depart from existing work in several respects. First, we develop alternative non-parametric FX volatility models, applying in particular a RNN architecture with a loop back from the output layer implying an error feedback mechanism, i.e. we apply a *non-linear error-correction* modelling approach to FX volatility.

Second, we apply our non-parametric models to *FX volatility*, something that has not been done so far. A recent development in the literature has been the application of non-parametric time series modelling approaches to volatility forecasts. Gaussian Kernel regression is an example, as in West and Cho (1995). Neural networks have also been found useful in modelling the properties of non-linear time series. As mentioned above, if there are quite a few articles on applications of NNR models to foreign exchange, stock and commodity markets,[1] there are rather few concerning financial markets volatility forecasting in general.[2] It seems therefore that, as an alternative technique to more traditional statistical forecasting methods, NNR models need further investigation to check whether they can add value in the field of foreign exchange volatility forecasting.

Finally, unlike previous work, we do not limit ourselves to forecasting accuracy but extend the analysis to the all important *trading efficiency*, taking advantage of the fact that there exists a large

[1] For NNR applications to commodity forecasting, see, for instance, Ntungo and Boyd (1998) and Trippi and Turban (1993). For applications to the stock market, see, among others, Deboeck (1994) and Leung *et al*. (2000).

[2] Even though there are no NNR applications yet to foreign exchange volatility forecasting, some researchers have used NNR models to measure the stock market volatility (see, for instance, Donaldson and Kamstra, 1997; Bartlmae and Rauscher, presentation at the FFM2000 Conference, London, 2000).

and liquid FX implied volatility market that enables us to apply sophisticated volatility trading strategies.

The paper is organized as follows. The next section describes our exchange rate and volatility data. The third section briefly presents the GARCH(1,1) model and gives the corresponding 21-day volatility forecasts. The fourth section provides a detailed overview and explains the procedures and methods used in applying the NNR and RNN modelling procedure to our financial time series, and it presents the 21-day volatility forecasts obtained with these methods. The fifth section briefly describes the model combinations retained and assesses the 21-day out-of-sample forecasts using traditional statistical accuracy criteria. The sixth section introduces the volatility trading strategy using FX option straddles that we follow once mispriced options have been identified through the use of our most successful volatility forecasting models. We present detailed trading results allowing for transaction costs and discuss their implications, particularly in terms of a qualified assessment of the efficiency of the currency options market. The final section provides some concluding comments and suggestions for further work.

THE EXCHANGE RATE AND VOLATILITY DATA

The motivation for this paper implies that the success or failure to develop profitable volatility trading strategies clearly depends on the possibility to generate accurate volatility forecasts and thus to implement adequate volatility modelling procedures.

Numerous studies have documented the fact that logarithmic returns of exchange rate time series exhibit 'volatility clustering' properties, that is, periods of large volatility tend to cluster together followed by periods of relatively lower volatility (see, among others, Baillie and Bollerslev, 1990; Jorion, 1997; Kroner *et al.*, 1995). Volatility forecasting crucially depends on identifying the typical characteristics of volatility within the restricted sample period selected and then projecting them over the forecasting period.

We present in turn the two databanks we have used for this study and the modifications to the original series we have made where appropriate.

The exchange rate series databank and historical volatility

The return series we use for the GBP/USD and USD/JPY exchange rates were extracted from a historical exchange rate database provided by Datastream. Logarithmic returns, defined as $\log(S_t/S_{t-1})$, are calculated for each exchange rate on a daily frequency basis. We multiply these returns by 100, so that we end up with percentage changes in the exchange rates considered, i.e. $s_t = 100.\log(S_t/S_{t-1})$.

Our exchange rate databank spans from 31 December 1993 to 9 May 2000, giving us 1610 observations per exchange rate.[3] This databank was divided into two separate sets with the first 1329 observations from 31 December 1993 to 9 April 1999 defined as our in-sample testing period and the remaining 280 observations from 12 April 1999 to 9 May 2000 being used for out-of-sample forecasting and validation.

In line with the findings of many earlier studies on exchange rate changes (see, among others, Baillie and Bollerslev, 1989; Engle and Bollerslev, 1986; Hsieh, 1989; West and Cho, 1995), the descriptive statistics of our currency returns (not reported here in order to conserve space) clearly

[3] In fact, we used exchange rate data from 1 November 1993 to 9 May 2000, the data during the period 1 November 1993 to 31 December 1993 being used for the 'pre-calculation' of the 21-day realized historical volatility.

show that they are nonnormally distributed and heavily fat-tailed. They also show that mean returns are not statistically different from zero. Further standard tests of autocorrelation, non-stationarity and heteroscedasticity show that logarithmic returns are all stationary and heteroscedastic. Whereas there is no evidence of autocorrelation for the GBP/USD return series, some autocorrelation is detected at the 10% significance level for USD/JPY returns.

The fact that our currency returns have zero unconditional mean enables us to use *squared returns* as a measure of their variance and *absolute returns* as a measure of their standard deviation or volatility.[4] The standard tests of autocorrelation, non-stationarity and heteroscedasticity (again not reported here in order to conserve space) show that squared and absolute currency returns series for the in-sample period are all non-normally distributed, stationary, autocorrelated and heteroscedastic (except USD/JPY squared returns which were found to be homoscedastic).

Still, as we are interested in analysing alternative volatility forecasting models and whether they can add value in terms of forecasting *realized* currency volatility, we must adjust our statistical computation of volatility to take into account the fact that, even if it is only the matter of a constant, in currency options markets, volatility is quoted in annualized terms. As we wish to focus on 1-month volatility forecasts and related trading strategies, taking, as is usual practice, a 252-trading day year (and consequently a 21-trading day month), we compute the 1-month volatility as the moving annualized standard deviation of our logarithmic returns and end up with the following historical volatility measures for the 1-month horizon:

$$\sigma_t = \frac{1}{21}\sum_{t-20}^{t}\left(\sqrt{252}\cdot|s_t|\right)$$

where $|s_t|$ is the absolute currency return.[5] The value σ_t is the realized 1-month exchange rate volatility that we are interested in forecasting as accurately as possible, in order to see if it is possible to find any mispriced option that we could possibly take advantage of.

The descriptive statistics of both historical volatility series (again not reported here in order to conserve space) show that they are non-normally distributed and fat-tailed. Further statistical tests of autocorrelation, heteroscedasticity and non-stationarity show that they exhibit strong autocorrelation but that they are stationary in levels. Whereas GBP/USD historical volatility is heteroscedastic, USD/JPY realized volatility was found to be homoscedastic.

Having presented our exchange rate series databank and explained how we compute our historical volatilities from these original series (so that they are in a format comparable to that which prevails in the currency options market), we now turn our attention to the implied volatility databank that we have used.

The implied volatility series databank

Volatility has now become an observable and traded quantity in financial markets, and particularly so in the currency markets. So far, most studies dealing with implied volatilities have used volatilities backed out from historical premium data on traded options rather than over-the-counter (OTC) volatility data (see, among others, Chiras and Manaster, 1978; Kroner *et al*., 1995; Latane and Rendleman, 1976; Lamoureux and Lastrapes, 1993; Xu and Taylor, 1996).

[4] Although the unconditional mean is zero, it is of course possible that the conditional mean may vary over time.

[5] The use of absolute returns (rather than their squared value) is justified by the fact that with zero unconditional mean, averaging absolute returns gives a measure of standard deviation.

As underlined by Dunis *et al*. (2000), the problem in using exchange data is that call and put prices are only available for given strike levels and fixed maturity dates. The corresponding implied volatility series must therefore be *backed out* using a specific option pricing model. This procedure generates two sorts of potential biases: material errors or mismatches can affect the variables that are needed for the solving of the pricing model, e.g. the forward points or the spot rate, and, more importantly, the very specification of the pricing model that is chosen can have a crucial impact on the final 'backed out' implied volatility series.

This is the reason why, in this paper, we use *data directly observable on the marketplace*. This original approach seems further warranted by current market practice whereby brokers and market makers in currency options deal in fact *in volatility terms* and not in option premium terms.[6] The volatility time series we use for the two exchange rates selected, GBP/USD and USD/JPY were extracted from a *market quoted implied volatilities* database provided by Chemical Bank for data until end-1996, and updated from Reuters 'Ric' codes subsequently. These at-the-money forward, market-quoted volatilities are in fact obtained from brokers by Reuters on a daily basis, at the close of business in London.

These implied volatility series are non-normally distributed and fat-tailed. Further statistical tests of autocorrelation and heteroscedasticity (again not reported here in order to conserve space) show that they exhibit strong autocorrelation and heteroscedasticity. Unit root tests show that, at the 1-month horizon, both GBP/USD and USD/JPY implied volatility are stationary at the 5% significance level.

Certainly, as noted by Dunis *et al*. (2001b) and confirmed in Tables A1.I and A1.III in Appendix 1 for the GBP/USD and USD/JPY, an interesting feature is that the mean level of implied volatilities stands well above average historical volatility levels.[7] This tendency of the currency options market to overestimate actual volatility is further documented by Figures A1.I and A1.II which show 1-month actual and implied volatilities for the GBP/USD and USD/JPY exchange rates. These two charts also clearly show that, for each exchange rate concerned, actual and implied volatilities are moving rather closely together, which is further confirmed by Tables A1.II and A1.IV for both GBP/USD and USD/JPY volatilities.

THE GARCH(1,1) BENCHMARK VOLATILITY FORECASTS

The choice of the benchmark model

As the GARCH model originally devised by Bollerslev (1986) and Taylor (1986) is well documented in the literature, we just present it very briefly, as it has now become widely used, in

[6] The market data that we use are *at-the-money forward volatilities*, as the use of either in-the-money or out-of-the-money volatilities would introduce a significant bias in our analysis due to the so-called 'smile effect', i.e. the fact that volatility is 'priced' higher for strike levels which are not at-the-money. It should be made clear that these implied volatilities are not simply backed out of an option pricing model but are instead directly quoted from brokers. Due to arbitrage they cannot diverge too far from the theoretical level.

[7] As noted by Dunis *et al*. (2001b), a possible explanation for implied volatility being higher than its historical counterpart may be due to the fact that market makers are generally options sellers (whereas end users are more often option buyers): there is probably a tendency among option writers to include a 'risk premium' when pricing volatility. Kroner *et al*. (1995) suggest another two reasons: (i) the fact that if interest rates are stochastic, then the implied volatility will capture both asset price volatility and interest rate volatility, thus skewing implied volatility upwards, and (ii) the fact that if volatility is stochastic but the option pricing formula is constant, then this additional source of volatility will be picked up by the implied volatility.

various forms, by both academics and practitioners to model conditional variance. We therefore do not intend to review its many different variants as this would be outside the scope of this paper. Besides, there is a wide consensus, certainly among market practitioners, but among many researchers as well that, when variants of the standard GARCH(1,1) model do provide an improvement, it is only marginal most of the time. Consequently, for this paper, we choose to estimate a GARCH(1,1) model for both the GBP/USD and USD/JPY exchange rates as it embodies a compact representation and serves well our purpose of finding an adequate *benchmark* for the more complex NNR models.

In its simple GARCH(1,1) form, the GARCH model basically states that the conditional variance of asset returns in any given period depends upon a constant, the previous period's squared random component of the return *and* the previous period's variance.

In other words, if we note σ_t^2 the conditional variance of the return at time t and ε_{t-1}^2 the squared random component of the return in the previous period, for a standard GARCH(1,1) process, we have:

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{1}$$

Equation (1) yields immediately the one-step-ahead volatility forecast and, using recursive substitution, Engle and Bollerslev (1986) and Baillie and Bollerslev (1992) give the n-step-ahead forecast for a GARCH(1,1) process:

$$\sigma_{t+n}^2 = \omega[1 + (\alpha+\beta) + \cdots + (\alpha+\beta)^{n-2}] + \omega + \alpha\varepsilon_t^2 + \beta\sigma_t^2 \tag{2}$$

This is the formula that we use to compute our GARCH(1,1) n-step-ahead out-of-sample forecast.

The GARCH(1,1) volatility forecasts

If many researchers have noted that no alternative GARCH specification could consistently outperform the standard GARCH(1,1) model, some as Bollerslev (1987), Baillie and Bollerslev (1989) and Hsieh (1989), among others, point out that the Student-t distribution fits the daily exchange rate logarithmic returns better than conditional normality, as the former is characterized by fatter tails. We thus generate GARCH(1,1) one-step-ahead forecasts with the Student-t distribution assumption.[8] We give our results for the GBP/USD exchange rate:

$$\begin{aligned}
\log(S_t/S_{t-1}) &= \varepsilon_t \\
\varepsilon_t|\varphi_{t-1} &\sim N(0,\ \sigma_t^2) \\
\sigma_t^2 &= \underset{(0.0015222)}{0.0021625} + \underset{(0.010135)}{0.032119}\ \varepsilon_{t-1}^2 + \underset{(0.013969)}{0.95864}\ \sigma_{t-1}^2
\end{aligned} \tag{3}$$

where the figures in parentheses are asymptotic standard errors. The t-values for α and β are highly significant and show strong evidence that σ_t^2 varies with ε_{t-1}^2 and σ_{t-1}^2. The coefficients also have

[8] Actually, we modelled conditional volatility with both the Normal and the t-distribution. The results are only slightly different. However, both the Akaike and the Schwarz Bayesian criteria tend to favour the t-distribution. We therefore selected the results from the t-distribution for further tests (see Appendix 2 for the USD/GBP detailed results).

the expected sign. Additionally, the conventional Wald statistic for testing the joint hypothesis that $\alpha = \beta = 0$ clearly rejects the null, suggesting a significant GARCH effect.

The parameters in equation (3) were used to estimate the 21-day-ahead volatility forecast for the USD/GBP exchange rate: using the one-step-ahead GARCH(1,1) coefficients, the conditional 21-day volatility forecast was generated each day according to equation (2) above. The same procedure was followed for the USD/JPY exchange rate volatility (see Appendix A2.III).

Figure 1 displays the GARCH(1,1) 21-day volatility forecasts for the USD/GBP exchange rate both in- and out-of-sample (the last 280 observations, from 12 April 1999 to 09 May 2000). It is clear that, overall, the GARCH model fits the realized volatility rather well during the in-sample period. However, during the out-of-sample period, the GARCH forecasts are quite disappointing. The USD/JPY out-of-sample GARCH(1,1) forecasts suffer from a similar inertia (see Figure A3.1 in Appendix 3).

In summary, if the GARCH(1,1) model can account for some statistical properties of daily exchange rate returns such as leptokurtosis and conditional heteroscedasticity, its ability to accurately predict volatility, despite its wide use among market professionals, is more debatable. In any case, as mentioned above, we only intend to use our GARCH(1,1) volatility forecasts as a benchmark for the non-linear non-parametric neural network models we intend to apply and test whether NNR/RNN models can produce a substantial improvement in the out-of-sample performance of our volatility forecasts.

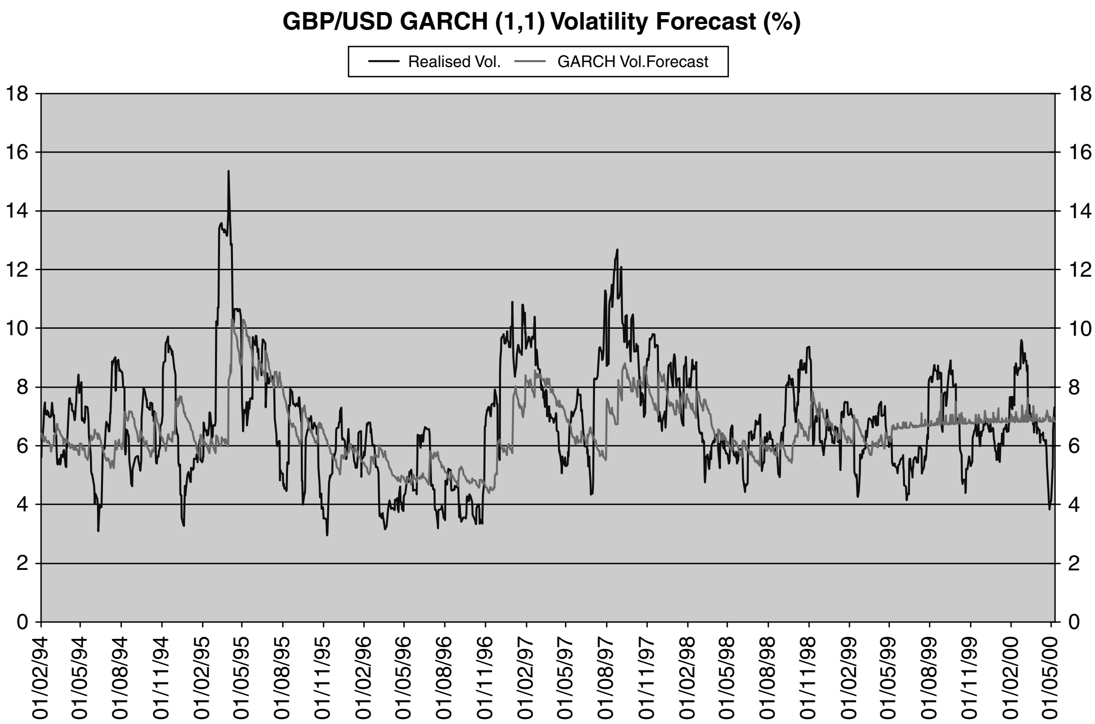

Figure 1. GBP/USD GARCH(1,1) volatility forecast

THE NEURAL NETWORK VOLATILITY FORECASTS

NNR modelling

Over the past few years, it has been argued that new technologies and quantitative systems based on the fact that most financial time series contain non-linearities have made traditional forecasting methods only second best. Neural Network Regression (NNR) models, in particular, have been applied with increasing success to economic and financial forecasting and would constitute the state of the art in forecasting methods (see, for instance, Zhang *et al.*, 1998).

It is clearly beyond the scope of this paper to give a complete overview of artificial neural networks, their biological foundation and their many architectures and potential applications (for more details, see, among others, Simpson, 1990; Hassoun, 1995).[9]

For our purpose, let it suffice to say that NNR models are a tool for determining the relative importance of an input (or a combination of inputs) for predicting a given outcome. They are a class of models made up of layers of elementary processing units, called neurons or nodes, which elaborate information by means of a non-linear transfer function. Most of the computing takes place in these processing units.

The input signals come from an input vector $A = (x^{[1]}, x^{[2]}, \ldots, x^{[n]})$ where $x^{[i]}$ is the activity level of the ith input. A series of weight vectors $W_j = (w_{1j}, w_{2j}, \ldots, w_{nj})$ is associated with the input vector so that the weight w_{ij} represents the strength of the connection between the input $x^{[i]}$ and the processing unit b_j. Each node may additionally have also a *bias input* θ_j modulated with the weight w_{0j} associated with the inputs. The total input of the node b_j is formally the dot product between the input vector A and the weight vector W_j, minus the weighted input bias. It is then passed through a non-linear transfer function to produce the output value of the processing unit b_j:

$$b_j = f\left(\sum_{i=1}^{n} x^{[i]} w_{ij} - w_{0j}\,\theta_j\right) = f(X_j) \tag{4}$$

In this paper, we have used the sigmoid function as activation function:[10]

$$f(X_j) = \frac{1}{1 + e^{-X_j}} \tag{5}$$

Figure 2 allows one to visualize a single output NNR model with one hidden layer and two hidden nodes, i.e. a model similar to those we developed for the GBP/USD and the USD/JPY volatility forecasts. The NNR model inputs at time t are $x_t^{[i]}(i = 1, 2, \ldots, 5)$. The hidden nodes outputs at time t are $h_t^{[j]}(j = 1, 2)$ and the NNR model output at time t is $\tilde{y}_t$, whereas the actual output is y_t.

At the beginning, the modelling process is initialized with random values for the weights. The output value of the processing unit b_j is then passed on to the single output node of the output

[9] In this paper, we use exclusively the multilayer perceptron, a multilayer feedforward network trained by error backpropagation.

[10] Other alternatives include the hyperbolic tangent, the bilogistic sigmoid, etc. A linear activation function is also a possibility, in which case the NNR model will be linear. Note that our choice of a sigmoid implies variations in the interval $]0, +1[$. Input data are thus normalized in the same range in order to present the learning algorithm with compatible values and avoid saturation problems.

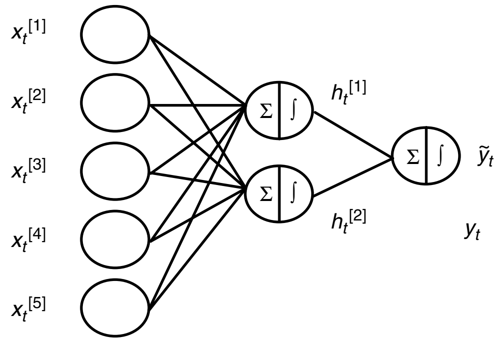

Figure 2. Single output NNR model

layer. The NNR error, i.e. the difference between the NNR forecast and the actual value, is analysed through the *root mean squared error*. The latter is systematically minimized by adjusting the weights according to the level of its derivative with respect to these weights. The adjustment obviously takes place in the direction that reduces the error.

As can be expected, NNR models with two hidden layers are more complex. In general, they are better suited for discontinuous functions; they tend to have better generalization capabilities but are also much harder to train. In summary, NNR model results depend crucially on the choice of the number of hidden layers, the number of nodes and the type of non-linear transfer function retained.

In fact, the use of NNR models further enlarges the forecaster's toolbox of available techniques by adding models where no specific functional form is *a priori* assumed.[11]

Following Cybenko (1989) and Hornik *et al.* (1989), it can be demonstrated that specific NNR models, if their hidden layer is sufficiently large, can approximate any continuous function.[12] Furthermore, it can be shown that NNR models are equivalent to *non-linear non-parametric models*, i.e. models where no decisive assumption about the generating process must be made in advance (see Cheng and Titterington, 1994).

Kouam *et al.* (1992) have shown that most forecasting models (ARMA models, bilinear models, autoregressive models with thresholds, non-parametric models with kernel regression, etc.) are embedded in NNR models. They show that each modelling procedure can in fact be written in the form of a network of neurons.

Theoretically, the advantage of NNR models over other forecasting methods can therefore be summarized as follows. As, in practice, the 'best' model for a given problem cannot be determined, it is best to resort to a modelling strategy which is a generalization of a large number of models, rather than to impose *a priori* a given model specification.

This has triggered an ever-increasing interest for applications to financial markets (see, for instance, Trippi and Turban, 1993; Deboeck, 1994; Rehkugler and Zimmermann, 1994; Refenes, 1995; Dunis, 1996).

Comparing NNR models with traditional econometric methods for foreign exchange rate forecasting has been the topic of several recent papers: Kuan and Liu (1995), Swanson and White

[11] Strictly speaking, the use of a NNR model implies assuming a functional form, namely that of the *transfer function*.
[12] This very feature also explains why it is so difficult to use NNR models, as one may in fact end up fitting the noise in the data rather than the underlying statistical process.

(1995) and Gençay (1996) show that NNR models can describe in-sample data rather well and that they also generate 'good' out-of-sample forecasts. Forecasting accuracy is usually defined in terms of small mean squared prediction error or in terms of directional accuracy of the forecasts. However, as mentioned already, there are still very few studies concerned with financial assets volatility forecasting.

RNN modelling

Recurrent neural network (RNN) models were introduced by Elman (1990). Their only difference from 'regular' NNR models is that they include a loop back from one layer, either the output or the intermediate layer, to the input layer. Depending on whether the loop back comes from the intermediate or the output layer, either the preceding values of the hidden nodes or the output error will be used as inputs in the next period. This feature, which seems welcome in the case of a forecasting exercise, comes at a cost: RNN models will require more connections than their NNR counterpart, thus accentuating a certain lack of transparency which is sometimes used to criticize these modelling approaches.

Using our previous notation and assuming the output layer is the one looped back, the RNN model output at time t depends on the inputs at time t and on the output at time $t-1$:[13]

$$\tilde{y}_t = F(x_t, \tilde{y}_{t-1}) \tag{6}$$

There is no theoretical answer as to whether one should preferably loop back the intermediate or the output layer. This is mostly an empirical question. Nevertheless, as looping back the output layer implies an error feedback mechanism, such RNN models can successfully be used for non-linear error-correction modelling, as advocated by Burgess and Refenes (1996). This is why we choose this particular architecture as an alternative modelling strategy for the GBP/USD and the USD/JPY volatility forecasts. Our choice seems further warranted by claims from Kuan and Liu (1995) and Tenti (1996) that RNN models are superior to NNR models when modelling exchange rates.

Figure 3 allows one to visualize a single-output RNN model with one hidden layer and two hidden nodes, again a model similar to those developed for the GBP/USD and the USD/JPY volatility forecasts.

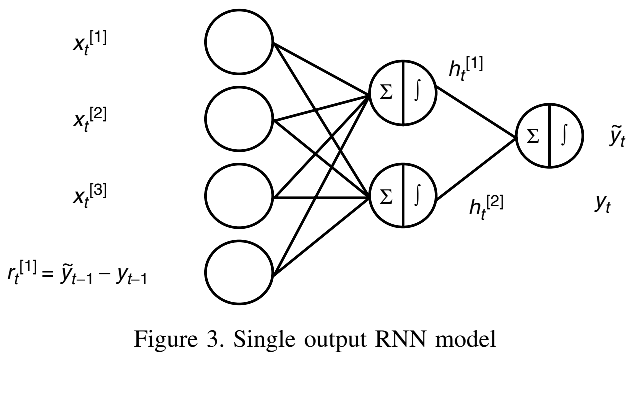

Figure 3. Single output RNN model

[13] With a loop back from the intermediate layer, the RNN output at time t depends on the inputs at time t and on the intermediate nodes at time $t-1$. Besides, the intermediate nodes at time t depend on the inputs at time t and on the hidden layer at time $t-1$. Using our notation, we have therefore: $\tilde{y}_t = F(x_t, h_{t-1})$, and $h_t = G(x_t, h_{t-1})$.

The NNR/RNN volatility forecasts

Input selection, data scaling and preprocessing

In the absence of an indisputable theory of exchange rate volatility, we assume that a specific exchange rate volatility can be explained by that rate's recent evolution, volatility spillovers from other financial markets, and macro-economic and monetary policy expectations.

In the circumstances, it seems reasonable to include, as potential inputs, exchange rate volatilities (including that which is to be modelled), the evolution of important stock and commodity prices, and, as a measure of macro-economic and monetary policy expectations, the evolution of the yield curve.[14]

As explained above (see footnote 10), all variables were normalized according to our choice of the sigmoid activation function. They had been previously transformed in logarithmic returns.[15]

Starting from a traditional linear correlation analysis, variable selection was achieved via a forward stepwise neural regression procedure: starting with both lagged historical and implied volatility levels, other potential input variables were progressively added, keeping the network architecture constant. If adding a new variable improved the level of explained variance over the previous 'best' model, the pool of explanatory variables was updated. If there was a failure to improve over the previous 'best' model after several attempts, variables in that model were alternated to check whether no better solution could be achieved. The model chosen finally was then kept for further tests and improvements.

Finally, conforming with standard heuristics, we partitioned our total data set into three subsets, using roughly 2/3 of the data for training the model, 1/6 for testing and the remaining 1/6 for validation. This partition in training, test and validation sets is made in order to control the error and reduce the risk of overfitting. Both the training and the following test period are used in the model-tuning process: the training set is used to develop the model; the test set measures how well the model interpolates over the training set and makes it possible to check during the adjustment whether the model remains valid for the future. As the fine-tuned system is not independent from the test set, the use of a third validation set which was not involved in the model's tuning is necessary. The validation set is thus used to estimate the actual performance of the model in a deployed environment.

In our case, the 1329 observations from 31 December 1993 to 9 April 1999 were considered as the in-sample period for the estimation of our GARCH(1,1) benchmark model. We therefore retain the first 1049 observations from 31 December 1993 to 13 March 1998 for the training set and the remainder of the in-sample period is used as test set. The last 280 observations from 12 April 1999 to 9 May 2000 constitute the validation set and serve as the out-of-sample forecasting period. This is consistent with the GARCH(1,1) model estimation.

Volatility forecasting results

We used two similar sets of input variables for the GBP/USD and USD/JPY volatilities, with the same output variable, i.e the realized 21-day volatility. Input variables included the lagged actual 21-day realized volatility ($Realized21_{t-21}$), the lagged implied 21-day volatility ($IVOL21_{t-21}$), lagged absolute logarithmic returns of the exchange rate ($|r|_{t-i}$, $i = 21, \ldots, 41$) and lagged logarithmic returns of the gold price ($DLGOLD_{t-i}$, $i = 21, \ldots, 41$) or of the oil price ($DLOIL_{t-i}$, $i = 21, \ldots, 41$), depending on the currency volatility being modelled.

[14] On the use of the yield curve as a predictor of future output growth and inflation, see, among others, Fama (1990) and Ivanova *et al.* (2000).

[15] Despite some contrary opinions, e.g. Balkin (presentation at the INFORMS Conference, Philadelphia, 1999), stationarity remains important if NNR/RNN models are to be assessed on the basis of the level of explained variance.

In terms of the final model selection, Tables A4.I(a) and A4.I(b) in Appendix 4 give the performance of the best NNR and RNN models over the validation (out-of-sample) data set for the USD/GBP volatility. For the same input space and architecture (i.e. with only one hidden layer), RNN models marginally outperform their NNR counterparts in terms of directional accuracy. This is important as trading profitability crucially depends on getting the direction of changes right. Tables A4.I(a) and A4.I(b) also compare models with only one hidden layer and models with two hidden layers while keeping the input and output variables unchanged: despite the fact that the best NNR model is a two hidden layer model with respectively 10 and 5 hidden nodes in each of its hidden layers, on average, NNR/RNN models with a single hidden layer perform marginally better while at the same time requiring less processing time.

The results of the NNR and RNN models for the USD/JPY volatility over the validation period are given in Tables A4.II(a) and A4.II(b) in Appendix 4. They are in line with those for the GBP/USD volatility, with RNN models outperforming their NNR counterparts and, in that case, the addition of a second hidden layer rather deteriorating performance.

Finally, we selected our two best NNR and RNN models for each volatility, NNR(44-10-5-1) and RNN(44-1-1) for the GBP/USD and NNR(44-1-1) and RNN(44-5-1) for the USD/JPY, to compare their out-of-sample forecasting performance with that of our GARCH(1,1) benchmark model. This evaluation is conducted on both statistical and financial criteria in the following sections. Yet, one can easily see from Figures A5.1 and A5.2 in Appendix 5 that, for both the GBP/USD and the USD/JPY volatilities, these out-of-sample forecasts do not suffer from the same degree of inertia as was the case for the GARCH(1,1) forecasts.

MODEL COMBINATIONS AND FORECASTING ACCURACY

Model combination

As noted by Dunis *et al*. (2001a), today most researchers would agree that individual forecasting models are misspecified in some dimensions and that the identity of the 'best' model changes over time. In this situation, it is likely that a combination of forecasts will perform better over time than forecasts generated by any individual model that is kept constant.

Accordingly, we build two rather simple model combinations to add to our three existing volatility forecasts, the GARCH(1,1), NNR and RNN forecasts.[16]

The simplest forecast combination method is the simple average of existing forecasts. As noted by Dunis *et al*. (2001b), it is often a hard benchmark to beat as other methods, such as regression-based methods, decision trees, etc., can suffer from a deterioration of their out-of-sample performance.

We call *COM1* the simple average of our GARCH(1,1), NNR and RNN volatility forecasts with the actual implied volatility (*IVOL21*). As we know, implied volatility is itself a popular method to measure market expectations of future volatility.

Another method of combining forecasts suggested by Granger and Ramanathan (1984) is to regress the in-sample historical 21-day volatility on the set of forecasts to obtain appropriate weights, and then apply these weights to the out-of-sample forecasts: it is noted *GR*. We follow Granger and Ramanathan's advice to add a constant term and not to constrain the weights to add to unity. We do not include both ANN and RNN forecasts in the regression as they can be highly collinear: for the USD/JPY, the correlation coefficient between both volatility forecasts is 0.984.

[16] More sophisticated combinations are possible, even based on NNR models as in Donaldson and Kamstra (1996), but this is beyond the scope of this paper.

We tried several alternative specifications for the Granger–Ramanathan approach. The parameters were estimated by ordinary least squares over the in-sample data set. Our best model for the GBP/USD volatility is presented below with t-statistics in parentheses, and the R-squared and standard error of the regression.

$$\underset{}{Actual_{t,21}} = \underset{(-6.550)}{-5.7442} + \underset{(7.712)}{0.7382\ \mathrm{RNN44}_{t,21}} + \underset{(8.592)}{0.6750\ \mathrm{GARCH}(1,1)_{t,21}} + \underset{(6.777)}{0.3226\ IVOL_{t,21}}$$

$$R^2 = 0.2805 \qquad \text{S.E. of regression} = 1.7129 \tag{7a}$$

For the USD/JPY volatility forecast combination, our best model was obtained using the NNR forecast rather than the RNN one:

$$Actual_{t,21} = \underset{(-7.091)}{-9.4293} + \underset{(7.561)}{1.5913\ \mathrm{NNR44}_{t,21}} + \underset{(0.975)}{0.0614\ \mathrm{GARCH}(1,1)_{t,21}} + \underset{(2.029)}{0.1701\ IVOL_{t,21}}$$

$$R^2 = 0.4128 \qquad \text{S.E. of regression} = 4.0239 \tag{7b}$$

As can be seen, the RNN/NNR-based forecast gets the highest weight in both cases, suggesting that the *GR* forecast relies more heavily on the RNN/NNR model forecasts than on the others. Figures A6.1 and A6.2 in Appendix 6 show that the *GR* and *COM1* forecast combinations, as the NNR and RNN forecasts, do not suffer from the same inertia as the GARCH(1,1) out-of-sample forecasts do.

We now have five volatility forecasts on top of the implied volatility 'market forecast' and proceed to test their out-of sample forecasting accuracy through traditional statistical criteria.

Out-of-sample forecasting accuracy

As is standard in the economic literature, we compute the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE) and Theil U-statistic (Theil-U). These measures have already been presented in detail by, among others, Makridakis *et al.* (1983), Pindyck and Rubinfeld (1998) and Theil (1966). We also compute a 'correct directional change' (CDC) measure which is described below.

Calling σ the actual volatility and $\hat{\sigma}$ the forecast volatility at time τ, with a forecast period going from $t+1$ to $t+n$, the forecast error statistics are respectively:

$$\mathrm{RMSE} = \sqrt{(1/n)\sum_{\tau=t+1}^{t+n}(\hat{\sigma}_\tau - \sigma_\tau)^2}$$

$$\mathrm{MAE} = (1/n)\sum_{\tau=t+1}^{t+n}\left|\hat{\sigma}_\tau - \sigma_\tau\right|$$

$$\text{Theil-}U = \sqrt{(1/n)\sum_{\tau=t+1}^{t+n}(\hat{\sigma}_\tau - \sigma_\tau)^2} \Bigg/ \left[\sqrt{(1/n)\sum_{\tau=t+1}^{t+n}\hat{\sigma}_\tau^2} + \sqrt{(1/n)\sum_{\tau=t+1}^{t+n}\sigma_\tau^2}\right]$$

$$\mathrm{CDC} = (100/n)\sum_{\tau=t+1}^{t+n} D_\tau$$

where $D_\tau = 1$ if $(\sigma_\tau - \sigma_{\tau-1})\cdot(\hat{\sigma}_\tau - \sigma_{\tau-1}) > 0$, else $D_\tau = 0$.

The RMSE and the MAE statistics are scale-dependent measures but give us a basis to compare our volatility forecasts with the realized volatility. The Theil-U statistics is independent of the scale of the variables and is constructed in such a way that it necessarily lies between zero and one, with zero indicating a perfect fit.

For all these three error statistics retained the lower the output, the better the forecasting accuracy of the model concerned. However, rather than on securing the lowest statistical forecast error, the profitability of a trading system critically depends on taking *the right position* and therefore getting the direction of changes right. RMSE, MAE and Theil-U are all important error measures, yet they may not constitute the best criterion from a profitability point of view. The CDC statistics is used to check whether the direction given by the forecast is the same as the actual change which has subsequently occurred and, for this measure, the higher the output the better the forecasting accuracy of the model concerned. Tables I and II compare, for the GBP/USD and the USD/JPY volatility respectively, our five volatility models and implied volatility in terms of the four accuracy measures retained.

These results are most interesting. Except for the GARCH(1,1) model (for all criteria for the USD/JPY volatility and in terms of directional change only for the GBP/USD volatility), they show that our five volatility forecasting models offer much more precise indications about future volatility than implied volatilities. This means that our volatility forecasts may be used to identify mispriced options, and a profitable trading rule can possibly be established based on the difference between the prevailing implied volatility and the volatility forecast.

The two NNR/RNN models and the two combination models predict correctly directional change at least over 59% of the time for the USD/JPY volatility.

Furthermore, for both volatilities, these models outperform the GARCH(1,1) benchmark model on all evaluation criteria. As a group, NNR/RNN models show superior out-of-sample forecasting performance on any statistical evaluation criterion, except directional change for which they are outperformed by model combinations. Within this latter group, the *GR* model performance is overall

Table I. GBP/USD Volatility models forecasting accuracy

GBP/USD Vol.	RMSE	MAE	Theil-U	CDC
IVOL21	1.98	1.63	0.13	49.64
GARCH(1,1)	1.70	1.48	0.12	48.57
NNR(44-10-5-1)	1.69	1.42	0.12	50.00
RNN(44-1-1)	1.50	1.27	0.11	52.86
COM1	1.65	1.41	0.11	65.23
GR	1.67	1.37	0.12	67.74

Table II. USD/JPY Volatility models forecasting accuracy

USD/JPY Vol.	RMSE	MAE	Theil-U	CDC
IVOL21	3.04	2.40	0.12	53.21
GARCH(1,1)	4.46	4.14	0.17	52.50
NNR(44-1-1)	2.41	1.88	0.10	59.64
RNN(44-5-1)	2.43	1.85	0.10	59.29
COM1	2.72	2.28	0.11	63.08
GR	2.70	2.13	0.11	66.67

the best in terms of statistical forecasting accuracy. The *GR* model combination provides the best forecast of directional change, achieving a remarkable directional forecasting accuracy of around 67% for the GBP/USD and the USD/JPY volatility.

Still, as noted by Dunis (1996), a good forecast may be a necessary but it is certainly not a sufficient condition for generating positive trading returns. Prediction accuracy is not the ultimate goal in itself and should not be used as the main guiding selection criterion for system traders. In the following section, we therefore use our volatility forecasting models to identify mispriced foreign exchange options and endeavour to develop profitable currency volatility trading models.

FOREIGN EXCHANGE VOLATILITY TRADING MODELS

Volatility trading strategies

Kroner *et al*. (1995) point out that, since expectations of future volatility play such a critical role in the determination of option prices, better forecasts of volatility should lead to a more accurate pricing and should therefore help an option trader to identify over- or underpriced options. Therefore a profitable trading strategy can be established based on the difference between the prevailing market implied volatility and the volatility forecast. Accordingly, Dunis and Gavridis (1997) advocate superimposing a volatility trading strategy on the volatility forecast.

As mentioned previously, there is a narrow relationship between volatility and the option price. An option embedding a high volatility gives the holder a greater chance of a more profitable exercise. When trading volatility, using at-the-money forward (ATMF) straddles, i.e. combining an ATFM call with an ATFM put with opposite deltas, results in taking no forward risk. Furthermore, as noted, among others, by Hull (1997), both the ATMF call and put have the same vega and gamma sensitivity. There is no directional bias.

If a large rise in volatility is predicted, the trader will buy both call and put. Although this will entail paying two premia, the trader will profit from a subsequent movement in volatility: if the foreign exchange market moves far enough either up or down, one of the options will end deeply in-the-money and, when it is sold back to the writing counterparty, the profit will more than cover the cost of both premia. The other option will expire worthless. Conversely, if both the call and put expire out-of-the-money following a period of stability in the foreign exchange market, only the premia will be lost.

If a large drop in volatility is predicted, the trader will sell the straddle and receive the two option premia. This a high-risk strategy if his market view is wrong as he might theoretically suffer unlimited loss, but, if he is right and both options expire worthless, he will have cashed in both premia.

The currency volatility trading models

The trading strategy adopted is based on the currency volatility trading model proposed by Dunis and Gavridis (1997). A long volatility position is initiated by buying the 1-month ATMF foreign exchange straddle if the 1-month volatility forecast is above the prevailing 1-month implied volatility level by more than a certain threshold used as a confirmation filter or reliability indicator. Conversely, a short ATMF straddle position is initiated if the 1-month volatility forecast is below the prevailing implied volatility level by more than the given threshold.

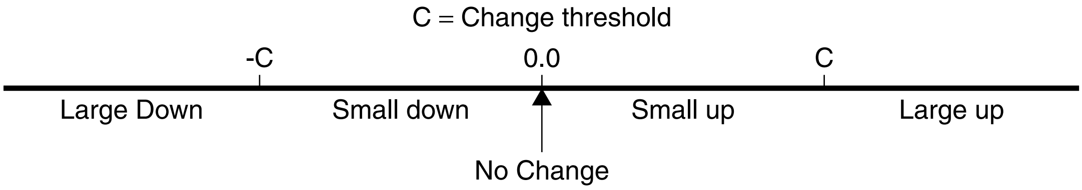

Figure 4. Volatility forecasts classification

To this effect, the first stage of the currency volatility trading strategy is, based on the threshold level as in Dunis (1996), to band the volatility predictions into 5 classes, namely, 'large up move', 'small up move', 'no change', 'large down move' and 'small down move' (Figure 4).

The change threshold defining the boundary between small and large movements was determined as a confirmation filter. Different strategies with filters ranging from 0.5 to 2.0 were analysed and are reported with our results.

The second stage is to decide the trading entry and exit rules. With our filter rule, a position is only initiated when the 1-month volatility forecast is above or below the prevailing 1-month implied volatility level by more than the threshold. That is:

- If $D_t > c$, then buy the ATMF straddle,
- If $D_t < -c$, then sell the ATFM straddle,

where D_t denotes the difference between the 1-month volatility forecast and the prevailing 1-month implied volatility, and c represents the threshold (or filter).

In terms of exit rules, our main test is to assume that the straddle is held until expiry and that no new positions can be initiated until the existing straddle has expired. As, due to the drop in time value during the life of an option, this is clearly not an optimal trading strategy, we also consider the case of American options which can be exercised at any time until expiry, and thus evaluate this second strategy assuming that positions are only held for five trading days (as opposed to one month).[17]

As in Dunis and Gavridis (1997), profitability is determined by comparing the level of implied volatility at the inception of the position with the prevailing 1-month realized historical volatility at maturity.

It is further weighted by the amount of the position taken, itself a function of the difference between the 1-month volatility forecast and the prevailing 1-month implied volatility level on the day when the position is initiated: intuitively, it makes sense to assume that, if we have a 'good' model, the larger $|D_t|$, the more confident we should be about taking the suggested position and the higher the expected profit. Calling G this gearing of position, we thus have:[18]

$$G = |D_t|/|c| \tag{8}$$

[17] For the 'weekly' trading strategy, we also considered closing out European options before expiry by taking the opposite position, unwinding positions at the prevailing implied volatility market rate after five trading days: this strategy was generally not profitable.

[18] Laws and Gidman (2000) adopt a similar strategy with a slightly different definition of the gearing.

Profitability is therefore defined as a volatility net profit (i.e. it is calculated in volatility points or '*vols*' as they are called by options traders[19]). Losses are also defined as a volatility loss, which implies two further assumptions: when short the straddle, no stop-loss strategy is actually implemented and the losing trade is closed out at the then prevailing volatility level (it is thus reasonable to assume that we overestimate potential losses in a real-world environment with proper risk management controls); when long the straddle, we approximate true losses by the difference between the level of implied volatility at inception with the prevailing volatility level when closing out the losing trade, whereas realized losses would only amount to the premium paid at the inception of the position (here again, we seem to overestimate potential losses). It is further assumed that volatility profits generated during one period are not reinvested during the next. Finally, in line with Dunis and Gavridis (1997), transaction costs of 25 b.p. per trade are included in our profit and loss computations.

Trading simulation results

The currency volatility trading strategy was applied from 31 December 1993 to 9 May 2000. Tables III and IV document our results for the GBP/USD and USD/JPY monthly trading strategies both for the in-sample period from 31 December 1993 to 9 April 1999 and the out-of-sample period from 12 April 1999 to 9 May 2000. The evaluation discussed below is focused on out-of-sample performance.

For our trading simulations, four different thresholds ranging from 0.5 to 2.0 and two different holding periods, i.e. monthly and weekly, have been retained. A higher threshold level implies requiring a higher degree of reliability in the signals and obviously reduces the overall number of trades.

The profitability criteria include the cumulative profit and loss with and without gearing, the total number of trades and the percentage of profitable trades. We also show the average gearing of the positions for each strategy.

First, we compare the performance of the NNR/RNN models with the benchmark GARCH(1,1) model. For the GBP/USD monthly volatility trading strategy in Table III, the GARCH(1,1) model generally produces higher cumulative profits not only in-sample but also out-of-sample. NNR/RNN models seldom produce a higher percentage of profitable trades in-sample or out-of-sample, although the geared cumulative return of the strategy based on the RNN(44-1-1) model is close to that produced with the benchmark model. With NNR/RNN models predicting more accurately directional change than the GARCH model, one would have intuitively expected them to show a better trading performance for the monthly volatility trading strategies.

This expected result is in fact achieved by the USD/JPY monthly volatility trading strategy, as shown in Table IV: NNR/RNN models clearly produce a higher percentage of profitable trades both in- and out-of-sample, with the best out-of-sample performance being that based on the RNN(44-5-1) model. On the contrary, the GARCH(1,1) model-based strategies produce very poor trading results, often recording an overall negative cumulative profit and loss figure.

Second, we evaluate the performance of model combinations. It is quite disappointing as, for both monthly volatility trading strategies, model combinations produce on average much lower

[19] In market jargon, 'vol' refers to both implied volatility and the measurement of volatility in per cent per annum (see, among others, Malz, 1996). Monetary returns could only be estimated by comparing the actual profit/loss of a straddle once closed out or expired against the premium paid/received at inception, an almost impossible task with OTC options.

Table III. GBP/USD Monthly volatility trading strategy

1 Threshold = 0.5 Trading days = 21

Sample observation period	In-sample (1–1329) (31/12/1993–09/04/1999)					Out-of-sample (1330–1610) (12/04/1999–09/05/2000)				
Models	NNR(44-10-5-1)	RNN(44-1-1)	GARCH(1,1)	COM1	GR	NNR(44-10-5-1)	RNN(44-1-1)	GARCH(1,1)	COM1	GR
P/L without gearing	58.39%	54.47%	81.75%	54.25%	81.40%	5.77%	5.22%	12.90%	10.35%	11.96%
P/L with gearing	240.44%	355.73%	378.18%	182.38%	210.45%	16.60%	35.30%	**42.91%**	16.41%	19.44%
Total trades	59	59	61	58	61	12	12	11	10	10
Profitable trades	67.80%	70.00%	77.05%	70.69%	83.61%	50.00%	58.33%	72.73%	70.00%	80.00%
Average gearing	2.83	4.05	3.87	2.35	2.13	1.55	2.17	2.39	1.46	1.44

2 Threshold = 1.0 Trading days = 21

Sample observation period	In-sample (1–1329) (31/12/1993–09/04/1999)					Out-of-sample (1330–1610) (12/04/1999–09/05/2000)				
Models	NNR(44-10-5-1)	RNN(44-1-1)	GARCH(1,1)	COM1	GR	NNR(44-10-5-1)	RNN(44-1-1)	GARCH(1,1)	COM1	GR
P/L without gearing	61.48%	57.52%	82.61%	64.50%	60.24%	7.09%	12.74%	12.08%	8.33%	9.65%
P/L with gearing	134.25%	190.80%	211.61%	116.99%	88.26%	8.65%	20.23%	**20.79%**	10.53%	11.03%
Total trades	51	51	58	45	47	7	8	9	5	6
Profitable trades	72.55%	69.09%	84.48%	80.00%	78.72%	85.71%	87.50%	66.67%	80.00%	83.33%
Average gearing	1.68	2.21	2.08	1.53	1.32	1.18	1.61	1.48	1.19	1.13

(*continued overleaf*)

Table III. (*Continued*)

3 Threshold = 1.5 Trading days = 21

Sample observation period	In-sample (1–1329) (31/12/1993–09/04/1999)					Out-of-sample (1330–1610) (12/04/1999–09/05/2000)				
Models	NNR(44-10-5-1)	RNN(44-1-1)	GARCH(1,1)	COM1	GR	NNR(44-10-5-1)	RNN(44-1-1)	GARCH(1,1)	COM1	GR
P/L without gearing	53.85%	61.13%	66.24%	62.26%	39.22%	9.26%	8.67%	8.93%	10.36%	8.69%
P/L with gearing	74.24%	114.88%	113.04%	80.65%	49.52%	11.16%	**11.75%**	11.32%	11.00%	9.92%
Total trades	40	40	52	31	24	4	6	6	3	2
Profitable trades	80.00%	71.43%	80.77%	83.87%	83.33%	100.00%	83.33%	83.33%	100.00%	100.00%
Average gearing	1.28	1.62	1.43	1.24	1.19	1.12	1.31	1.22	1.05	1.14

4 Threshold = 2.0 Trading days = 21

Sample observation period	In-sample (1–1329) (31/12/1993–09/04/1999)					Out-of-sample (1330–1610) (12/04/1999–09/05/2000)				
Models	NNR(44-10-5-1)	RNN(44-1-1)	GARCH(1,1)	COM1	GR	NNR(44-10-5-1)	RNN(44-1-1)	GARCH(1,1)	COM1	GR
P/L without gearing	69.03%	63.04%	60.57%	48.35%	20.97%	4.39%	7.80%	10.54%	4.39%	—
P/L with gearing	103.33%	98.22%	85.19%	63.05%	24.25%	5.37%	5.71%	**11.29%**	4.70%	—
Total trades	24	33	31	16	8	1	4	4	1	0
Profitable trades	82.76%	78.57%	79.49%	94.12%	88.89%	100.00%	100.00%	100.00%	100.00%	—
Average gearing	1.35	1.40	1.24	1.25	1.16	1.22	1.09	1.06	1.07	—

Note: Cumulative P/L figures are expressed in volatility points.

Table IV. USD/JPY Monthly volatility trading strategy

1 Threshold = 0.5 Trading days = 21

Sample observation Period	In-sample (1–1329) (31/12/1993–09/04/1999)					Out-of-sample (1330–1610) (12/04/1999–09/05/2000)				
Models	NNR(44-1-1)	RNN(44-5-1)	GARCH(1,1)	COM1	GR	NNR(44-1-1)	RNN(44-5-1)	GARCH(1,1)	COM1	GR
P/L without gearing	31.35%	16.42%	−26.42%	19.15%	19.73%	16.21%	20.71%	−8.57%	16.19%	3.21%
P/L with gearing	151.79%	144.52%	6.93%	152.36%	82.92%	63.61%	**106.17%**	−4.42%	75.78%	11.50%
Total trades	62	62	60	60	61	13	13	12	12	13
Profitable trades	54.84%	51.61%	38.33%	53.33%	60.66%	76.92%	84.62%	50.00%	66.67%	61.54%
Average gearing	3.89	3.98	3.77	2.41	2.36	2.86	3.91	5.73	2.63	1.86

2 Threshold = 1.0 Trading days = 21

Sample observation Period	In-sample (1–1329) (31/12/1993–09/04/1999)					Out-of-sample (1330–1610) (12/04/1999–09/05/2000)				
Models	NNR(44-1-1)	RNN(44-5-1)	GARCH(1,1)	COM1	GR	NNR(44-1-1)	RNN(44-5-1)	GARCH(1,1)	COM1	GR
P/L without gearing	25.83%	44.59%	−9.71%	40.60%	64.35%	20.70%	21.32%	−1.94%	13.53%	26.96%
P/L with gearing	67.66%	105.01%	43.09%	76.72%	122.80%	**52.81%**	45.81%	42.14%	21.41%	46.53%
Total trades	58	58	57	51	52	12	12	12	11	12
Profitable trades	62.07%	56.90%	36.84%	58.82%	69.23%	83.33%	83.33%	41.67%	63.64%	83.33%
Average gearing	2.01	2.16	1.97	1.58	1.55	1.97	1.94	3.03	1.41	1.66

(*continued overleaf*)

Table IV. (*Continued*)

3 Threshold = 1.5 Trading days = 21

Sample observation Period	In-sample (1–1329) (31/12/1993–09/04/1999)					Out-of-sample (1330–1610) (12/04/1999–09/05/2000)				
Models	NNR(44-1-1)	RNN(44-5-1)	GARCH(1,1)	COM1	GR	NNR(44-1-1)	RNN(44-5-1)	GARCH(1,1)	COM1	GR
P/L without gearing	47.07%	23.62%	40.93%	46.63%	84.17%	19.65%	25.54%	1.87%	5.09%	23.80%
P/L with gearing	92.67%	75.69%	86.60%	75.70%	109.49%	40.49%	**73.19%**	31.31%	10.65%	32.91%
Total trades	51	51	46	37	42	10	10	12	6	10
Profitable trades	64.71%	55.77%	52.17%	59.46%	73.81%	80.00%	80.00%	33.33%	50.00%	90.00%
Average gearing	1.71	1.72	1.60	1.33	1.35	1.54	1.94	2.21	1.37	1.41

4 Threshold = 2.0 Trading days = 21

Sample observation Period	In-sample (1–1329) (31/12/1993–09/04/1999)					Out-of-sample (1330–1610) (12/04/1999–09/05/2000)				
Models	NNR(44-1-1)	RNN(44-5-1)	GARCH(1,1)	COM1	GR	NNR(44-1-1)	RNN(44-5-1)	GARCH(1,1)	COM1	GR
P/L without gearing	52.52%	28.98%	39.69%	27.89%	75.99%	34.35%	33.70%	0.11%	7.74%	3.09%
P/L with gearing	202.23%	72.21%	49.77%	35.77%	94.30%	60.33%	**64.76%**	−5.11%	12.35%	5.48%
Total trades	37	37	41	21	26	10	10	12	3	7
Profitable trades	59.46%	61.54%	56.10%	61.90%	80.77%	90.00%	90.00%	33.00%	100.00%	71.00%
Average gearing	1.67	1.59	1.39	1.25	1.26	1.54	1.68	1.54	1.54	1.23

Note: Cumulative P/L figures are expressed in volatility points.

cumulative returns than alternative strategies based on NNR/RNN models for the USD/JPY volatility and on either the GARCH(1,1) or the RNN(44-1-1) model for the GBP/USD volatility. As a general rule, the *GR* combination model fails to clearly outperform the simple average model combination *COM1* during the out-of-sample period, something already noted by Dunis *et al*. (2001b).

Overall, with the monthly holding period, RNN model-based strategies show the strongest out-of-sample trading performance: in terms of geared cumulative profit, they come first in four out of the eight monthly strategies analysed, and second best in the remaining four cases. The strategy with the highest return yields a 106.17% cumulative profit over the out-of-sample period and is achieved for the USD/JPY volatility with the RNN(44-5-1) model and a filter equal to 0.5.

The results of the weekly trading strategy are presented in Tables A7.I and A7.II in Appendix 7. They basically confirm the superior performance achieved through the use of RNN model-based strategies and the comparatively weak results obtained through the use of model combination.

Finally, allowing for transaction costs, it is worth noting that all the trading strategies retained produce positive returns, except some based on the GARCH(1,1) benchmark model for the USD/JPY volatility. RNN models appear as the best single modelling approach for short-term volatility trading. Somewhat surprisingly, model combination, the overall best performing approach in terms of forecasting accuracy, fails to improve the RNN-based volatility trading results.

CONCLUDING REMARKS AND FURTHER WORK

The rationale for this paper was to develop a non-linear non-parametric approach to forecast FX volatility, identify mispriced options and subsequently develop a trading strategy based upon this modelling procedure.

Using daily data from December 1993 through April 1999, we examined the use of Neural Network Regression (NNR) and Recurrent Neural Network (RNN) regression models for *forecasting* and subsequently *trading* currency volatility, with an application to the GBP/USD and USD/JPY exchange rates.

These models were then tested *out-of-sample* over the period April 1999–May 2000, not only in terms of *forecasting accuracy*, but also in terms of *trading performance*. In order to do so, we applied a realistic volatility trading strategy using FX option straddles once mispriced options had been identified.

Allowing for transaction costs, most of the trading strategies retained produced positive returns. RNN models appeared as the best single modelling approach in a short-term trading context.

Model combination, despite its superior performance in terms of forecasting accuracy, failed to produce superior trading strategies. Admittedly, other combination procedures such as decision trees, neural networks, as in Donaldson and Kamstra (1996), and unanimity or majority voting schemes as applied by Albanis and Batchelor (2001) should be investigated.

Further work is also needed to compare the results from NNR and RNN models with those from more 'refined' parametric or semiparametric approaches than our GARCH(1,1) benchmark model, such as Donaldson and Kamstra (1997), So *et al*. (1999), Bollen *et al*. (2000), Flôres and Roche (2000) and Beine and Laurent (2001).

Finally, applying dynamic risk management, the trading strategy retained could also be refined to integrate more realistic trading assumptions than those of a fixed holding period of either 5 or 21 trading days.

However, despite the limitations of this paper, we were clearly able to develop reasonably accurate FX volatility forecasts, identify mispriced options and subsequently simulate a profitable trading strategy. In the circumstances, the unambiguous implication from our results is that, for the period and currencies considered, the currency option market was inefficient and/or the pricing formulae applied by market participants were inadequate.

APPENDIX 1

Table A1.I. Summary statistics—GBP/USD realized and implied 1-month volatility (31/12/1993–09/04/1999)

Sample observations 1 to 1329		
Variable(s)	Historical vol.	Implied vol.
Maximum	15.3644	15.0000
Minimum	2.9448	3.2500
Mean	6.8560	8.2575
Std deviation	1.9928	1.6869
Skewness	0.69788	0.38498
Kurtosis—3	0.81838	1.1778
Coef of variation	0.29067	0.20429

Table A1.II. Correlation matrix of realized and implied volatility (GBP/USD)

	Realized vol.	Implied vol.
Realized vol.	1.0000	0.79174
Implied vol.	0.79174	1.0000

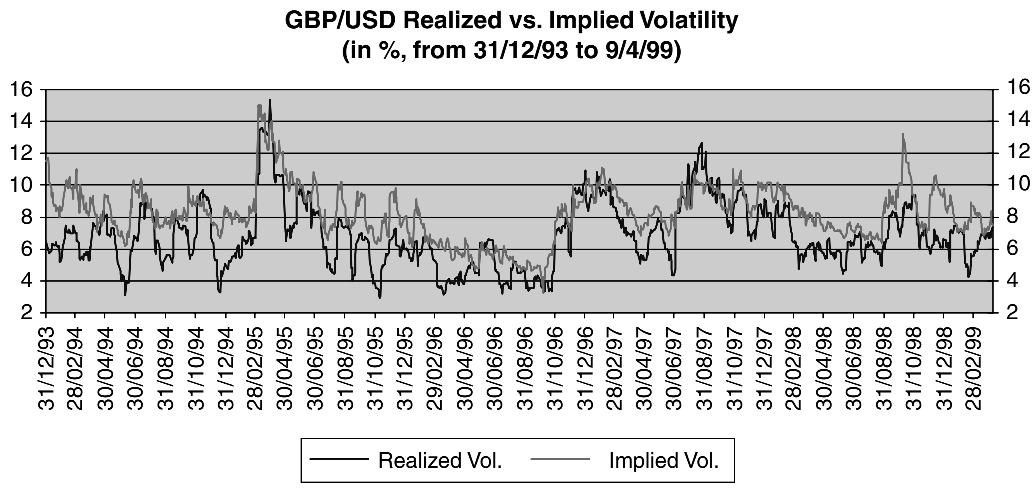

Figure A1.1. GBP/USD realized and implied volatility

Table A1.III. Summary statistics—USD/JPY realized and implied 1-month volatility (31/12/1993–09/04/1999)

Sample observations 1 to 1329		
Variable(s)	Historical vol.	Implied vol.
Maximum	33.0446	35.0000
Minimum	4.5446	6.1500
Mean	11.6584	12.2492
Std deviation	5.2638	3.6212
Skewness	1.4675	0.91397
Kurtosis—3	2.6061	2.1571
Coef of variation	0.45150	0.29563

Table A1.IV. Correlation matrix of realized and implied volatility (USD/JPY)

	Realized vol.	Implied vol.
Realized vol.	1.0000	0.80851
Implied vol.	0.80851	1.0000

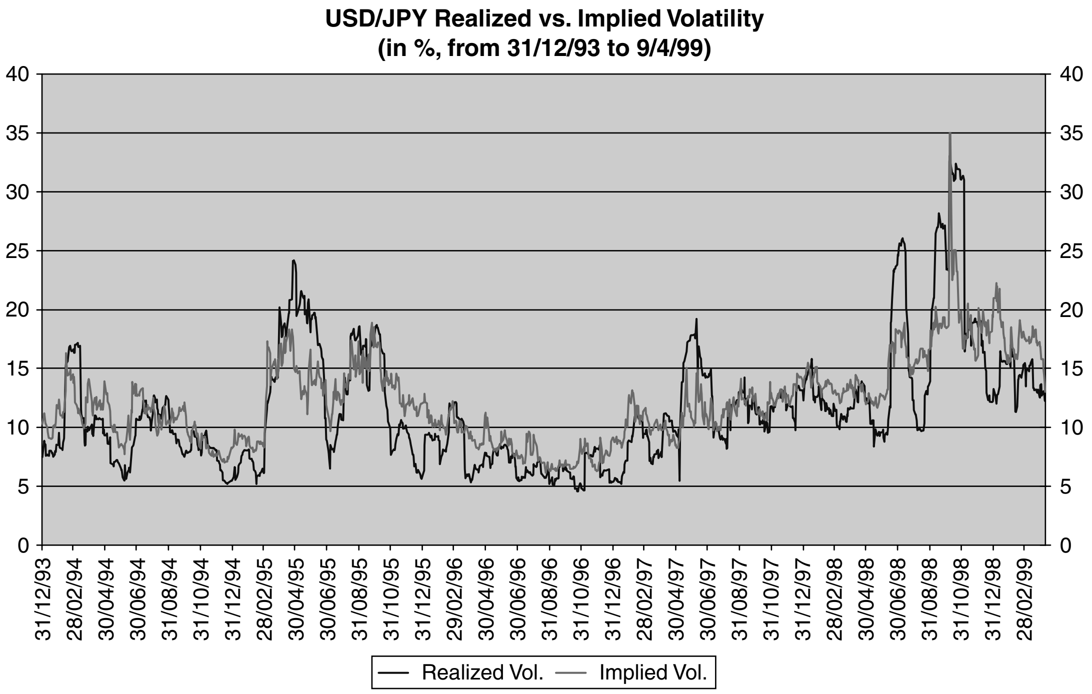

Figure A1.2. USD/JPY realized and implied volatility

APPENDIX 2

Table A2.I. GBP/USD GARCH (1,1) assuming a t-distribution and Wald test

GBP/USD GARCH(1,1) assuming a t-distribution converged after 30 iterations

Dependent variable is DLUSD (1327 observations used for estimation from 3 to 1329)			
Regressor	Coefficient	Standard error	T-Ratio[Prob]
ONE	0.0058756	0.010593	0.55466[0.579]
DLUSD(−1)	0.024310	0.027389	0.88760[0.375]
R-squared	0.0011320	R-Bar-Squared	−0.0011330
S.E. of regression	0.45031	F-stat. F(3,1323)	0.49977[0.682]
Mean of dependent variable	0.0037569	S.D. of dependent variable	0.45006
Residual sum of squares	268.2795	Equation log-likelihood	−749.6257
Akaike info. criterion	−754.6257	Schwarz Bayesian criterion	−767.6024
DW-statistic	1.9739		
Parameters of the conditional heteroscedastic model explaining H-SQ, the conditional variance of the error term			
	Coefficient	Asymptotic standard error	
Constant	0.0021625	0.0015222	
E-SQ(−1)	0.032119	0.010135	
H-SQ(−1)	0.95864	0.013969	
D.F. of t-Dist.	5.1209	0.72992	

H-SQ stands for the conditional variance of the error term.
E-SQ stands for the square of the error term.

Wald test of restriction(s) imposed on parameters

Based on GARCH regression of DLUSD on: ONE DLUSD(−1) 1327 observations used for estimation from 3 to 1329	
Coefficients A1 to A2 are assigned to the above regressors respectively. Coeffs. B1 to B4 are assigned to ARCH parameters respectively List of restriction(s) for the Wald test: b2 = 0; b3 = 0	
Wald statistic	CHSQ(2) = 16951.2[0.000]

Table A2.II. GBP/USD GARCH (1,1) assuming a Normal distribution and Wald test

GBP/USD GARCH(1,1) assuming a Normal distribution converged after 35 literations

Dependent variable is DLUSD (1327 observations used for estimation from 3 to 1329)			
Regressor	Coefficient	Standard error	T-Ratio[Prob]
ONE	0.0046751	0.011789	0.39657[0.692]
DLUSD(−1)	0.047043	0.028546	1.6480[0.100]
R-squared	0.0011651	R-Bar-Squared	−0.0010999
S.E. of regression	0.45030	F-stat. F(3,1323)	0.51439[0.672]
Mean of dependent variable	0.0037569	S.D. of dependent variable	0.45006
Residual sum of squares	268.2707	Equation log-likelihood	−796.3501
Akaike info. criterion	−800.3501	Schwarz Bayesian criterion	−810.7315
DW-statistic	2.0199		

Table A2.II. (*Continued*)

Parameters of the conditional heteroscedastic model explaining H-SQ, the conditional variance of the error term		
	Coefficient	Asymptotic standard error
Constant	0.0033874	0.0016061
E-SQ(−1)	0.028396	0.0074743
H-SQ(−1)	0.95513	0.012932

H-SQ stands for the conditional variance of the error term.
E-SQ stands for the square of the error term.

Wald test of restriction(s) imposed on parameters

Based on GARCH regression of DLUSD on: ONE DLUSD(−1) 1327 observations used for estimation from 3 to 1329	
Coefficients A1 to A2 are assigned to the above regressors respectively. Coeffs. B1 to B3 are assigned to ARCH parameters respectively List of restriction(s) for the Wald test: b2 = 0; b3 = 0	
Wald statistic	CHSQ(2) = 16941.1[0.000]

Table A2.III. USD/JPY GARCH (1,1) assuming a t-distribution and Wald test

USD/JPY GARCH(1,1) assuming a t-distribution converged after 26 iterations

Dependent variable is DLYUSD (1327 observations used for estimation from 3 to 1329)			
Regressor	Coefficient	Standard error	T-Ratio[Prob]
ONE	0.037339	0.015902	2.3480[0.019]
DLYUSD(−1)	0.022399	0.026976	0.83032[0.407]
R-squared	0.0010778	R-Bar-Squared	−0.0011874
S.E. of regression	0.79922	F-stat. $F(3,1323)$	0.47580[0.699]
Mean of dependent variable	0.0056665	S.D. of dependent variable	0.79875
Residual sum of squares	845.0744	Equation Log-likelihood	−1385.3
Akaike info. criterion	−1390.3	Schwarz Bayesian criterion	−1403.3
DW-statistic	1.8999		

Parameters of the conditional heteroscedastic model explaining H-SQ, the conditional variance of the error term		
	Coefficient	Asymptotic standard error
Constant	0.0078293	0.0045717
E-SQ(−1)	0.068118	0.021094
H-SQ(−1)	0.92447	0.023505
D.F. of t-Dist.	4.3764	0.54777

H-SQ stands for the conditional variance of the error term.
E-SQ stands for the square of the error term.

Table A2.III. (*Continued*)

Wald test of restriction(s) imposed on parameters

Based on GARCH regression of DLYUSD on: ONE DLYUSD(−1) 1327 observations used for estimation from 3 to 1329	
Coefficients A1 to A2 are assigned to the above regressors respectively. Coeffs. B1 to B4 are assigned to ARCH parameters respectively List of restriction(s) for the Wald test: b2 = 0; b3 = 0	
Wald statistic	CHSQ(2) = 11921.3[0.000]

APPENDIX 3

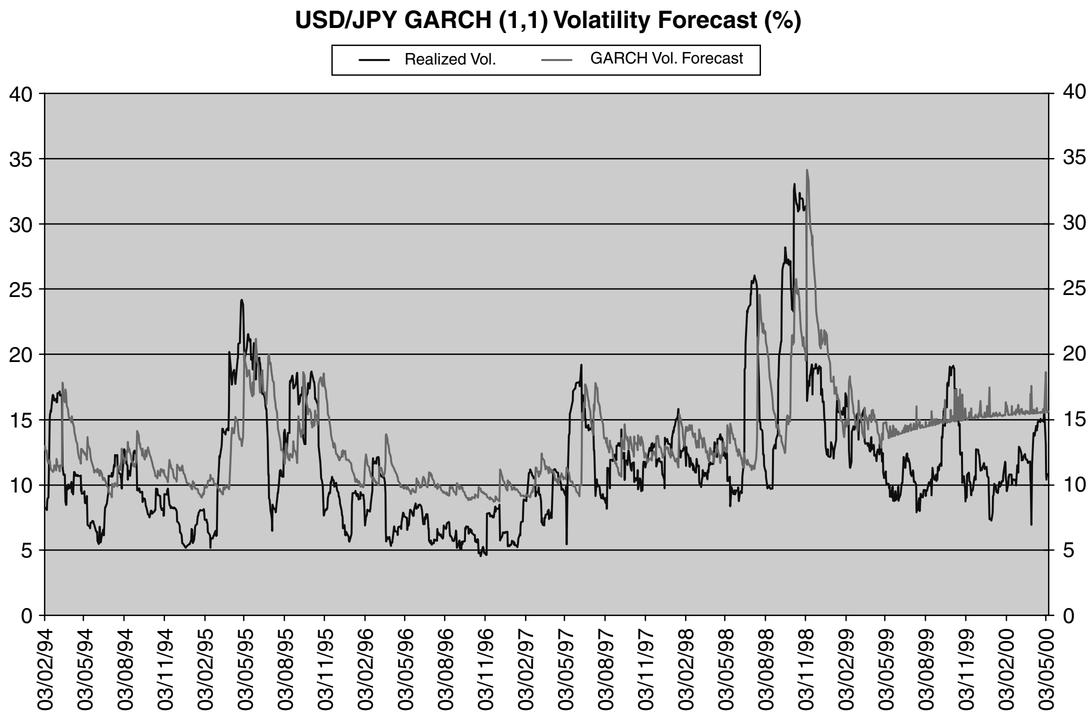

Figure A3.1. USD/JPY GARCH(1,1) volatility forecast

APPENDIX 4

Table A4.I. GBP/USD NNR and RNN test results

(a) GBP/USD NNR test results for the validation data set

	NNR(44-1-1)	NNR(44-5-1)	NNR(44-10-1)	**NNR(44-10-5-1)**	NNR(44-15-10-1)
Explained variance	1.4%	5.9%	8.1%	12.5%	15.6%
Average relative error	0.20	0.20	0.20	0.20	0.21
Average absolute error	1.37	1.39	1.40	1.41	1.44
Average direction error	33.3%	32.3%	31.5%	30.5%	31.5%

(b) GBP/USD RNN test results for the validation data set

	RNN(44-1-1)	RNN(44-5-1)	RNN(44-10-1)	RNN(44-10-5-1)	RNN(44-15-10-1)
Explained variance	13.3%	9.3%	5.5%	6.8%	12.0%
Average relative error	0.18	0.19	0.19	0.20	0.20
Average absolute error	1.25	1.29	1.33	1.38	1.42
Average direction error	30.1%	32.6%	32.6%	30.8%	32.3%

NNR/RNN (*a-b-c*) represents different neural network models, where:
a = number of input variables
b = number of hidden nodes
c = number of output nodes
$Realized\_Vol\ (t) = f[IVol(t-21), Realized\_Vol\ (t-21), |r|(t-21, \ldots, t-41), DLGOLD\ (t-21, \ldots, t-41)]$

Table A4.II. USD/JPY NNR and RNN test results

(a) USD/JPY NNR test results for the validation data set

	NNR(44-1-1)	NNR(44-5-1)	NNR(44-10-1)	NNR(44-10-5-1)	NNR(44-15-10-1)
Explained variance	5.1%	5.4%	5.4%	2.5%	2.9%
Average relative error	0.16	0.16	0.16	0.16	0.16
Average absolute error	1.88	1.87	1.86	1.85	1.84
Average direction error	30.1%	30.8%	30.8%	32.6%	32.3%

(b) USD/JPY RNN test results for the validation data set

	RNN(44-1-1)	**RNN(44-5-1)**	RNN(44-10-1)	RNN(44-10-5-1)	RNN(44-15-10-1)
Explained variance	8.4%	8.5%	8.0%	3.2%	3.0%
Average relative error	0.16	0.16	0.16	0.16	0.16
Average absolute error	1.86	1.85	1.84	1.85	1.85
Average direction error	30.1%	29.4%	29.4%	31.5%	30.1%

NNR/RNN (*a-b-c*) represents different neural network models, where:
a = number of input variables
b = number of hidden nodes
c = number of output nodes
$Realized\_Vol\ (t) = f[IVol(t-21), Realized\_Vol\ (t-21), |r|(t-21, \ldots, t-41), DLOIL\ (t-21, \ldots, t-41)]$

APPENDIX 5

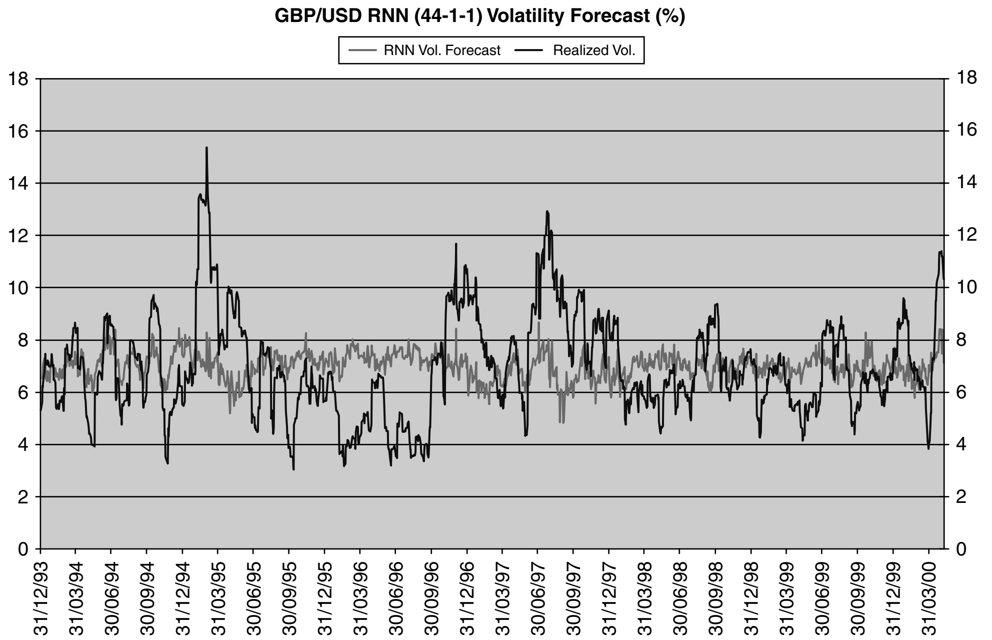

Figure A5.1. GBP/USD RNN(44-1-1) volatility forecast

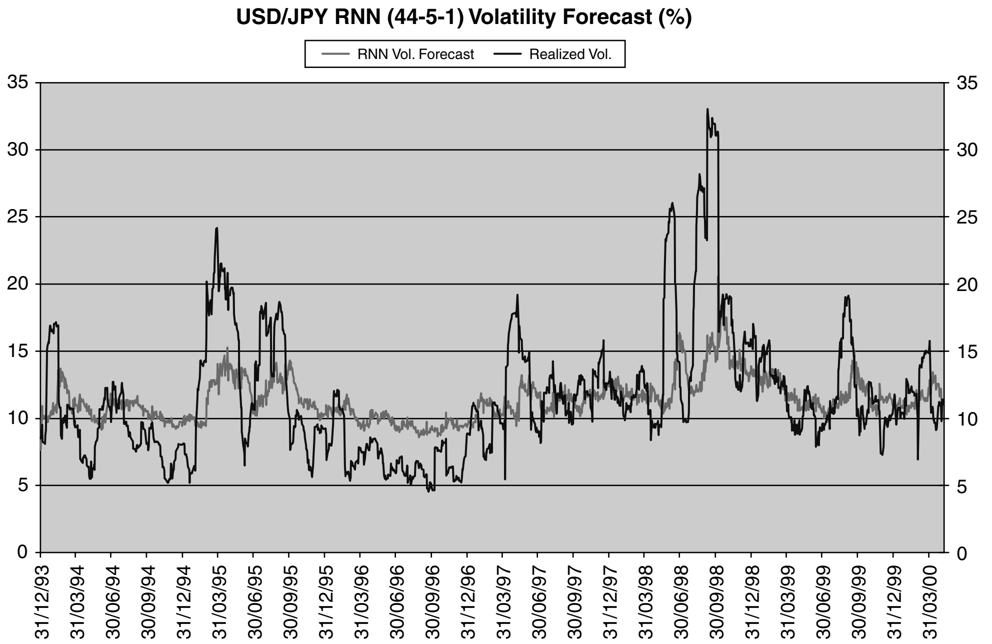

Figure A5.2. USD/JPY RNN(44-5-1) volatility forecast

APPENDIX 6

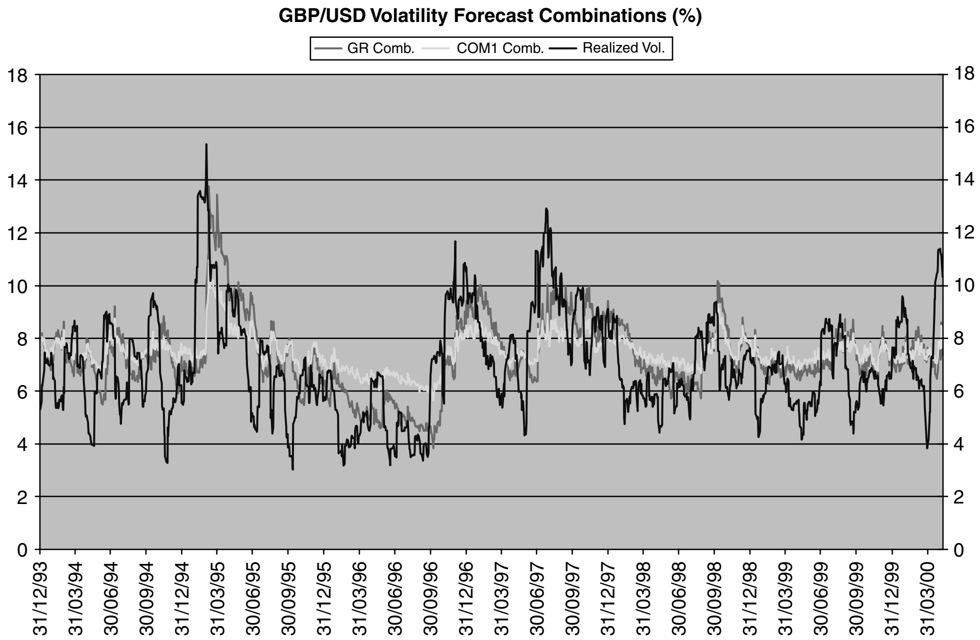

Figure A6.1. GBP/USD volatility forecast combinations

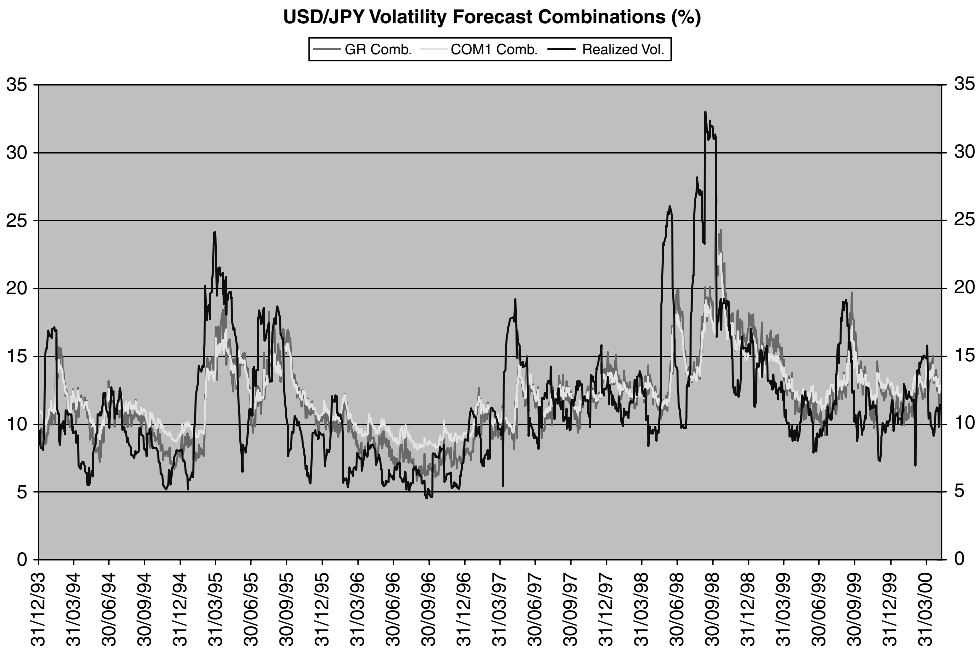

Figure A6.2. USD/JPY volatility forecast combinations

APPENDIX 7

Table A7.I. GBP/USD Weekly volatility trading strategy

1 Threshold = 0.5 Trading days = 5

Sample observation period	In-sample (1–1329) (31/12/1993–09/04/1999)					Out-of-sample (1330–1610) (12/04/1999–09/05/2000)				
Models	NNR(44-10-5-1)	RNN(44-1-1)	GARCH(1,1)	COM1	GR	NNR(44-10-5-1)	RNN(44-1-1)	GARCH(1,1)	COM1	GR
P/L without gearing	162.15%	160.00%	262.41%	189.82%	313.55%	15.37%	13.30%	−12.85%	8.30%	30.08%
P/L with gearing	626.62%	852.70%	889.04%	597.16%	792.87%	36.46%	**58.98%**	−2.39%	20.92%	53.64%
Total trades	221	221	224	210	232	39	43	40	35	33
Profitable trades	66.97%	68.07%	81.70%	72.86%	91.38%	53.85%	60.47%	30.00%	45.71%	72.73%
Average gearing	2.86	4.20	2.71	2.54	2.19	1.72	2.58	1.82	1.62	1.63

2 Threshold = 1.0 Trading days = 5

Sample observation period	In-sample (1–1329) (31/12/1993–09/04/1999)					Out-of-sample (1330–1610) (12/04/1999–09/05/2000)				
Models	NNR(44-10-5-1)	RNN(44-1-1)	GARCH(1,1)	COM1	GR	NNR(44-10-5-1)	RNN(44-1-1)	GARCH(1,1)	COM1	GR
P/L without gearing	166.15%	147.13%	213.13%	167.03%	243.71%	20.09%	12.02%	7.87%	11.11%	18.57%
P/L with gearing	333.20%	420.39%	434.96%	300.56%	366.44%	**27.01%**	20.87%	12.80%	14.27%	23.42%
Total trades	166	166	153	147	138	21	27	17	13	14
Profitable trades	73.49%	72.00%	84.31%	77.55%	94.93%	76.19%	55.56%	58.82%	76.92%	85.71%
Average gearing	1.82	2.35	1.73	1.61	1.42	1.24	1.64	1.31	1.16	1.21

3 Threshold = 1.5 Trading days = 5

Sample observation period	In-sample (1–1329) (31/12/1993–09/04/1999)					Out-of-sample (1330–1610) (12/04/1999–09/05/2000)				
Models	NNR(44-10-5-1)	RNN(44-1-1)	GARCH(1,1)	COM1	GR	NNR(44-10-5-1)	RNN(44-1-1)	GARCH(1,1)	COM1	GR
P/L without gearing	128.90%	137.62%	176.30%	122.22%	129.73%	16.49%	7.08%	7.08%	7.08%	4.39%
P/L with gearing	198.76%	278.91%	268.10%	173.28%	158.83%	**19.49%**	10.52%	8.56%	7.61%	5.21%
Total trades	113	113	94	82	61	10	18	4	4	2
Profitable trades	74.34%	69.70%	93.62%	82.93%	95.08%	90.00%	50.00%	100.00%	100.00%	100.00%
Average gearing	1.49	1.81	1.45	1.38	1.21	1.12	1.27	1.19	1.07	1.14

4 Threshold = 2.0 Trading days = 5

Sample observation period	In-sample (1–1329) (31/12/1993–09/04/1999)					Out-of-sample (1330–1610) (12/04/1999–09/05/2000)				
Models	NNR(44-10-5-1)	RNN(44-1-1)	GARCH(1,1)	COM1	GR	NNR(44-10-5-1)	RNN(44-1-1)	GARCH(1,1)	COM1	GR
P/L without gearing	95.08%	115.09%	118.43%	62.47%	46.67%	3.26%	7.25%	4.67%	3.26%	—
P/L with gearing	131.34%	194.33%	158.35%	79.39%	52.97%	3.99%	**9.27%**	5.65%	3.48%	—
Total trades	68	68	56	36	18	1	10	2	1	0
Profitable trades	82.35%	72.44%	94.64%	88.89%	94.44%	100.00%	80.00%	100.00%	100.00%	—
Average gearing	1.34	1.55	1.31	1.29	1.14	1.22	1.14	1.16	1.07	—

Note: Cumulative P/L figures are expressed in volatility points.

Table A7.II. USD/JPY Weekly volatility trading strategy

1 Threshold = 0.5 Trading days = 5

Sample observation Period	In-sample (1–1329) (31/12/1993–09/04/1999)					Out-of-sample (1330–1610) (12/04/1999–09/05/2000)				
Models	NNR(44-1-1)	RNN(44-5-1)	GARCH(1,1)	COM1	GR	NNR(44-1-1)	RNN(44-5-1)	GARCH(1,1)	COM1	GR
P/L without gearing	−10.17%	−11.66%	208.83%	77.06%	387.38%	63.95%	78.77%	−82.07%	39.23%	69.31%
P/L with gearing	−79.31%	−32.08%	1675.67%	517.50%	1095.73%	251.14%	**372.70%**	−359.44%	111.86%	206.43%
Total trades	251	251	242	232	229	50	50	53	48	50
Profitable trades	51.00%	50.60%	54.13%	55.17%	78.60%	76.00%	86.00%	18.87%	60.42%	78.00%
Average gearing	3.76	3.88	3.67	2.67	2.42	3.22	3.73	5.92	2.18	2.35

2 Threshold = 1.0 Trading days = 5

Sample observation Period	In-sample (1–1329) (31/12/1993–09/04/1999)					Out-of-sample (1330–1610) (12/04/1999–09/05/2000)				
Models	NNR(44-1-1)	RNN(44-5-1)	GARCH(1,1)	COM1	GR	NNR(44-1-1)	RNN(44-5-1)	GARCH(1,1)	COM1	GR
P/L without gearing	18.73%	30.43%	234.66%	142.37%	308.50%	52.70%	63.01%	−75.61%	31.71%	76.88%
P/L with gearing	0.43%	102.50%	897.30%	331.50%	578.10%	88.19%	**135.49%**	−162.97%	49.88%	124.63%
Total trades	213	213	201	168	165	39	43	53	30	37
Profitable trades	52.11%	54.17%	54.23%	58.93%	82.42%	76.92%	81.40%	24.53%	66.67%	94.59%
Average gearing	2.16	2.16	2.15	1.68	1.65	1.77	1.98	2.96	1.52	1.56

Table A7.II. (*Continued*)

3 Threshold = 1.5 Trading days = 5

Sample observation Period	In-sample (1–1329) (31/12/1993–09/04/1999)					Out-of-sample (1330–1610) (12/04/1999–09/05/2000)				
Models	NNR(44-1-1)	RNN(44-5-1)	GARCH(1,1)	COM1	GR	NNR(44-1-1)	RNN(44-5-1)	GARCH(1,1)	COM1	GR
P/L without gearing	34.15%	54.46%	251.07%	104.67%	282.96%	64.00%	55.83%	−65.02%	5.33%	52.59%
P/L with gearing	60.28%	164.86%	643.47%	203.58%	399.48%	**108.52%**	81.09%	−102.88%	11.99%	77.99%
Total trades	157	157	144	98	104	33	35	50	13	22
Profitable trades	54.78%	56.21%	63.89%	60.20%	86.54%	81.82%	80.00%	24.00%	69.00%	100.00%
Average gearing	1.83%	1.75	1.78	1.45	1.35	1.56	1.51	2.00	1.38	1.42

4 Threshold = 2.0 Trading days = 5

Sample observation Period	In-sample (1–1329) (31/12/1993–09/04/1999)					Out-of-sample (1330–1610) (12/04/1999–09/05/2000)				
Models	NNR(44-1-1)	RNN(44-5-1)	GARCH(1,1)	COM1	GR	NNR(44-1-1)	RNN(44-5-1)	GARCH(1,1)	COM1	GR
P/L without gearing	−6.13%	14.08%	247.22%	71.30%	154.88%	49.61%	53.90%	−43.71%	9.23%	30.33%
P/L with gearing	−35.86%	63.43%	514.30%	105.49%	202.86%	71.66%	**86.20%**	−52.64%	14.94%	40.28%
Total trades	117	117	113	47	50	21	23	47	6	11
Profitable trades	51.28%	52.94%	70.80%	65.96%	86.00%	90.48%	86.96%	28.00%	83.00%	100.00%
Average gearing	1.58	1.66	1.56	1.28	1.25	1.46	1.52	1.74	1.55	1.30

Note: Cumulative P/L figures are expressed in volatility points.

ACKNOWLEDGMENTS

We are grateful to Professor Ken Holden and Professor John Thompson of Liverpool Business School and to an anonymous referee for helpful comments on an earlier version of this paper. The usual disclaimer applies. The opinions expressed herein are not those of Girobank.

REFERENCES

Albanis GT, Batchelor RA. 2001. 21 nonlinear ways to beat the market. In *Developments in Forecast Combination and Portfolio Choice*, Dunis C, Moody J, Timmermann A (eds). John Wiley: Chichester.

Baillie RT, Bollerslev T. 1989. The message in daily exchange rates: a conditional variance tale. *Journal of Business and Economic Statistics* **7**: 297–305.

Baillie RT, Bollerslev T. 1990. Intra-day and inter-market volatility in foreign exchange rates. *Review of Economic Studies* **58**: 565–585.

Baillie RT, Bollerslev T. 1992. Prediction in dynamic models with time-dependent conditional variances. *Journal of Econometrics* **52**: 91–113.

Beine M, Laurent S. 2001. Structural change and long memory in volatility: new evidence from daily exchange rates. In *Developments in Forecast Combination and Portfolio Choice*, Dunis C, Moody J, Timmermann A (eds). John Wiley: Chichester.

Black F, Scholes M. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* **81**: 637–654.

Bollen PB, Gray SF, Whaley RE. 2000. Regime-switching in foreign exchange rates: evidence from currency option prices. *Journal of Econometrics* **94**: 239–276.

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

Bollerslev T. 1987. A conditional heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* **69**: 542–547.

Burgess AN, Refenes AN. 1996. The use of error feedback terms in neural network modelling of financial time series. In *Forecasting Financial Markets*, Dunis C (ed.). John Wiley: Chichester.

Cheng B, Titterington DM. 1994. Neural networks: a review from a statistical perspective. *Statistical Science* **9**: 2–54.

Chiras DP, Manaster S. 1978. The information content of option prices and a test of market efficiency. *Journal of Financial Economics* **6**: 213–234.

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

Dave R, Stahl G. 1997. On the accuracy of VaR estimates based on the variance–covariance approach. Working Paper, Olsen and Associates, Zürich.

Deboeck GJ. 1994. *Trading on the Edge - Neural, Genetic and Fuzzy Systems for Chaotic Financial Markets*. John Wiley: New York.

Donaldson RG, Kamstra M. 1996. Forecast combining with neural networks. *Journal of Forecasting* **15**: 49–61.

Donaldson RG, Kamstra M. 1997. An artificial neural network–GARCH model for international stock return volatility. *Journal of Empirical Finance* **4**: 17–46.

Dunis C. 1996. The economic value of neural network systems for exchange rate forecasting. *Neural Network World* **1**: 43–55.

Dunis C, Gavridis M. 1997. Volatility trading models: an application to daily exchange rates. *Working Papers in Financial Economics*, BNP Global Markets Research **1**: 1–4.

Dunis C, Laws J, Chauvin S. 2000. FX volatility forecasts: a fusion-optimisation approach. *Neural Network World* **10**: 187–202.

Dunis C, Moody J, Timmermann A (eds). 2001a. *Developments in Forecast Combination and Portfolio Choice*. John Wiley: Chichester.

Dunis C, Laws J, Chauvin S. 2001b. The Use of market data and model combination to improve forecasting accuracy. In *Developments in Forecast Combination and Portfolio Choice*, Dunis C, Moody J, Timmermann A (eds). John Wiley: Chichester.

Elman JL. 1990. Finding structures in time. *Cognitive Sciences* **14**: 179–211.

Engle RF, Bollerslev T. 1986. Modelling the persistence of conditional variances. *Econometric Reviews* **5**: 1–50.

Engle RF, Hong CH, Kane A, Noh J. 1993. Arbitrage valuation of variance forecasts with simulated options. In *Advances in Futures and Options Research*, Chance DM, Tripp RR (eds). JIA Press: Greenwich, CT.

Fama EF. 1990. Term structure forecasts of interest rates, inflation and real returns. *Journal of Monetary Economics* **25**: 59–76.

Flôres RG Jr, Roche BB. 2000. Volatility modelling in the forex market: an empirical evaluation. In *Advances in Quantitative Asset Management*, Dunis C (ed.). Kluwer: Boston, MA.

Franses PH, Van Homelen P. 1998. On forecasting exchange rates using neural networks. *Applied Financial Economics* **8**: 589–596.

Garman MB, Kohlhagen SW. 1983. Foreign currency option values. *Journal of International Money and Finance* **2**: 231–237.

Gençay R. 1996. Non-linear prediction of security returns with moving average rules. *Journal of Forecasting* **15**: 165–174.

Granger CW, Ramanathan R. 1984. Improved methods of combining forecasts. *Journal of Forecasting* **3**: 197–204.

Hassoun MH. 1995. *Fundamentals of Artificial Neural Networks*. MIT Press: Cambridge, MA.

Hornik K, Stinchcombe M, White H. 1989. Multilayer feedforward networks are universal approximators. *Neural Networks* **2**: 359–366.

Hsieh DA. 1989. Modeling heteroscedasticity in daily foreign exchange rates. *Journal of Business and Economic Statistic* **7**: 307–317.

Hull JC. 1997. *Options, Futures and Other Derivative Securities*. Prentice Hall: Englewood Cliffs, NJ.

Ivanova D, Lahiri K, Seitz F. 2000. Interest rate spreads as predictors of German inflation and business cycles. *International Journal of Forecasting* **16**: 39–58.

Jorion P. 1995. Predicting volatility in the foreign exchange market. *Journal of Finance* **50**: 507–528.

Jorion P. 1997. *Value at Risk: The New Benchmark for Controlling Market Risk*. McGraw-Hill: New York.

Kouam A, Badran F, Thiria S. 1992. Approche méthodologique pour l'étude de la prévision à l'aide de réseaux de neurones. *Actes de Neuro-Nîmes.*

Kroner KF, Kneafsey KP, Claessens S. 1995. Forecasting volatility in commodity markets. *Journal of Forecasting* **14**: 77–95.

Kuan CM, Liu T. 1995. Forecasting exchange rates using feedforward and recurrent neural networks. *Journal of Applied Economics* **10**: 347–364.

Lamoureux CG, Lastrapes WD. 1993. Forecasting stock-return variances: toward an understanding of stochastic implied volatilities. *Review of Financial Studies* **6**: 293–326.

Latane HA, Rendleman RJ. 1976. Standard deviations of stock price ratios implied in option prices. *Journal of Finance* **31**: 369–381.

Laws J, Gidman A. 2000. Forecasting stock market volatility and the application of volatility trading models. *CIBEF Research Paper*, Liverpool Business School.

Leung MT, Daouk H, Chen A-S. 2000. Forecasting stock indices: a comparison of classification and level estimation models. *International Journal of Forecasting* **16**: 173–190.

Makridakis S, Wheelwright SC, McGee VE. 1983. *Forecasting: Methods and Applications*. John Wiley: New York.

Malz AM. 1996. Using option prices to estimate realignment probabilities in the European monetary system. In *Forecasting Financial Markets*, Dunis C (ed.). John Wiley: Chichester.

Nelson M, Hill T, Remus W, O'Connor M. 1999. Time series forecasting using neural networks: should the data be deseasonalized first? *Journal of Forecasting* **18**: 359–367.

Ntungo C, Boyd M. 1998. Commodity futures trading performance using neural network models versus ARIMA models. *Journal of Futures Markets* **8**: 965–983.

Pindyck RS, Rubinfeld DL. 1998. *Econometric Models and Economic Forecasts*. McGraw-Hill: New York.

Refenes AN. 1995. *Neural Networks In the Capital Markets*. John Wiley: Chichester.

Rehkugler H, Zimmermann HG (eds). 1994. *Neuronale Netze in der Ökonomie—Grundlagen und finanzwirtschaftliche Anwendungen*. Verlag Franz Vahlen: München.
RiskMetrics. 1994. *Technical Document*. Morgan Guaranty Trust Company: New York.
Simpson PK. 1990. *Artificial Neural Systems—Foundations, Paradigms, Applications, and Implementations*. Pergamon Press: New York.
So MKP, Lam K, Li WK. 1999. Forecasting exchange rate volatility using autoregressive random variance model. *Applied Financial Economics* **9**: 583–591.
Swanson NR, White H. 1995. A model-selection approach to assessing the information in the term-structure using linear models and neural networks. *Journal of Business and Economic Statistics* **13**: 265–275.
Taylor SJ. 1986. *Modelling Financial Time Series*. John Wiley: Chichester.
Tenti P. 1996. Forecasting foreign exchange rates using recurrent neural networks. *Applied Artificial Intelligence* **10**: 567–581.
Theil H. 1966. *Applied Economic Forecasting*. North-Holland: Amsterdam.
Trippi R, Turban E (eds). 1993. *Neural Networks in Finance and Investing—Using Artificial Intelligence to Improve Real-Word Performance*. Probus: Chicago.
West KD, Cho D. 1995. The predictive ability of several models of exchange rate volatility. *Journal of Econometrics* **69**: 367–391.
Xu X, Taylor SJ. 1996. Conditional volatility and the informational efficiency of the PHLX currency options market. In *Forecasting Financial Markets*, Dunis C (ed.). John Wiley: Chichester.
Zhang G, Patuwo BE, Hu MY. 1998. Forecasting with artificial neural networks: the state of the art. *International Journal of Forecasting* **14**: 35–62.

Authors' biographies:

Christian Dunis is Girobank Professor of Banking and Finance at Liverpool Business School and Director of its Centre for International Banking, Economics and Finance, CIBEF. He is also a Visiting Professor of International Finance at Venice International University. Current research interests include non-linear modelling, model combinations and quantitative trading.

Xuehuan Huang is an Associate Researcher with CIBEF. She graduated from Liverpool Business School with an MSc in International Banking and Finance and currently works as a financial analyst with Bayer.

Authors' address:

Christian Dunis and **Xuehuan Huang**, Centre for International Banking, Economics and Finance, JMU, John Foster Building, 98 Mount Pleasant, Liverpool L3 5UZ, UK.

