Alternative volatility models for risk management and trading: Application to the EUR/USD and USD/JPY rates

Christian L. Dunis and Yao Xian Chen*
*CIBEF — Centre for International Banking, Economics and Finance, JMU, John Foster Building, 98 Mount Pleasant, Liverpool L3 5UZ, UK. Tel: +44 (0151) 231 3867; E-mail: yxc119@hotmail.com

Received (in revised form): 20th October, 2004

Christian L. Dunis is Professor of Banking and Finance at Liverpool John Moores University, and Director of its Centre for International Banking, Economics and Finance (CIBEF). He is also a consultant to asset management firms on applications to finance of emerging software technologies. He is an Editor of the *European Journal of Finance* and has published widely in the field of financial markets analysis and forecasting.

Yao Xian Chen is an Associate Researcher at CIBEF. She holds an MSc in International Banking and Finance from the School of Accounting, Finance and Economics at Liverpool John Moores University.

Practical applications

The foreign exchange market is by far the largest financial market in the world. According to the last Bank for International Settlements triennial survey, the EUR/USD and USD/JPY exchange rates are the most heavily traded exchange rates representing some 45 per cent of the $1.9 trillion daily trading volume of the world currency markets. This paper focuses on these two heavily traded exchange rates, analysing the predictive power of alternative forecasting models of foreign exchange volatility from both a statistical and an economic point of view, the latter integrating both dimensions of trading and risk management. It also investigates whether implied volatility data obtained from the currency options market can add value in terms of forecasting accuracy: because there will never be such thing as unanimous agreement on the future volatility estimate, market participants with a better view of the evolution of volatility will have an edge over their competitors. In practice, those investors or market participants who can reliably predict volatility should be better able to control the financial risks and, at the same time, profit from their superior forecasting ability.

Abstract

This paper examines the forecasting ability of several alternative models of currency volatility applied to two foreign exchange rates: EUR/USD and USD/JPY which, according to the Bank for International Settlements (BIS), represent 45 per cent of the $1.9 trillion daily trading volume on the world currency markets.

Derivatives Use, Trading & Regulation, Vol. 11 No. 2, 2005, pp. 126–156

1747–4426

Benchmarked against two naïve 'random walk' models and a RiskMetrics volatility model, the predictive abilities of the autoregressive (AR(p)); generalised autoregressive conditional heteroscedasticity (GARCH(p,q)); new modelling approaches such as stochastic variance (SV) and neural network regression (NNR) models; and two different model combinations are assessed at the one-day, five-day and 21-day horizons not only in terms of traditional forecasting accuracy measures but also in terms of risk management efficiency under the value-at-risk (VaR) framework and trading performance with a volatility filter strategy. These daily models are developed for the period from 2nd January, 1998, to 13th May, 2002 (1,116 observations) and tested out-of-sample from 14th May, 2002 to 28th March, 2003 (223 observations). The essence of the contribution is three 'forecasting' competitions using the same forecasts, some obtained from new modelling techniques, for three different purposes: the first is statistical accuracy, the second VaR and the third is simulated trading. Although no single volatility model emerges as an overall winner in terms of forecasting accuracy, risk management efficiency and FX trading performance, 'mixed' models incorporating market data for currency volatility, NNR models and combinations of models perform best most of the time.

INTRODUCTION

Foreign exchange (FX) volatility has been a constant feature of the International Monetary System ever since the breakdown of the Bretton Woods system of fixed parities of 1971–1973. Not surprisingly, in the wake of the growing use of derivatives in other financial markets, and following the extension of the Black–Scholes option pricing model to foreign exchange by Garman and Kohlhagen,[1] 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–Kohlhagen pricing formula is precisely the future foreign exchange rate volatility during the life of the option, but 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. Simply because there will never be such a thing as 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.

Higher volatility implies a greater possible dispersion of the foreign exchange rate at expiry; all other things being equal, logically the option holder has an asset with a greater chance of a more profitable exercise. In practice, those investors/market participants who can reliably predict volatility should be better able to control the financial risks associated with their option positions and, at the same time, profit from their superior forecasting ability. Volatility forecasts are also useful in the management of risk, eg for putting together option hedging programmes, assessing value-at-risk (VaR), etc; hence the interest in volatility forecasting in the risk management literature. As that literature has matured, and as our abilities in computation and simulation have advanced, it has fuelled the development of powerful risk management methods and software.

Admittedly, FX volatility series show strong heteroscedasticity and non-linearity features, making their forecasting a truly demanding task.[2] A revolution in modelling and forecasting volatility began some two decades ago with Engle.[3] Since then, many different modelling approaches have been applied to volatility forecasting. With the exception of Engle *et al.*,[4] Laws and Gidman[5] and Dunis and Huang,[6] however, those papers evaluate the out-of-sample forecasting performance of their models using traditional statistical accuracy criteria, such as root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), Theil-U statistic (Theil-U) and correct directional change (CDC) 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 standard statistical criterion. Similarly, risk managers are more concerned with VaR precision than with volatility forecasting accuracy as such.

Accordingly, the motivation for this paper is to investigate the predictive power of alternative forecasting models of FX volatility, from both a statistical and an economic point of view — the latter integrating both dimensions of trading and risk management. Where new modelling techniques, such as non-linear non-parametric neural network regression (NNR) and the time-varying parameter stochastic variance (SV) models, have been applied to this field, their results have been gauged in terms of statistical accuracy. The essence of this contribution is to show three 'forecasting' competitions using the same forecasts for three different purposes: the first is statistical accuracy, the second VaR and the third is simulated trading.

The use of new modelling approaches such as NNR and SV models, is examined in this context of FX trading and risk management. The results of the NNR and SV models are benchmarked against two naïve 'random walk' models; RiskMetrics volatility; the simpler AR(p) and GARCH(p,q) models; and two model combinations: in terms of model combination, a simple average combination and the Granger/Ramanathan[7] optimal weighting regression-based approach are employed and their results investigated.

These 'pure' time series models are complemented with implied volatility data obtained from the FX options market, leading to the estimation of 'mixed' time series models. This implied volatility term is then examined to see whether it adds value in terms of forecasting accuracy and the trading and risk management applications retained.

As both volatility trading and risk management encompass short-term and more medium-term risks (respectively, trading risk and credit risk) necessitating both short-term and medium-term volatility forecasts, the focus is on one-day, one-week and one-month out forecasts (respectively one-, five- and 21-trading day horizons).

Using daily data for the two most heavily traded exchange rates according to the Bank for International Settlements (BIS)[8] triennial survey of the world currency markets, the EUR/USD and USD/JPY, the volatility models are developed for the period from 2nd January, 1998 to 13th May, 2002, and are tested out-of-sample from 14th May, 2002 to 28th March, 2003. The models are tested not only in terms of forecasting accuracy, but also, more importantly, in terms of risk management efficiency under the VaR framework and trading performance with the implementation of a volatility filter in a spot trading simulation.

The empirical results clearly show that while, in terms of statistical accuracy, NNR models and the naïve historical volatility (HVOL) benchmark model perform as the best single modelling techniques, AR models based on squared returns seem to work best in terms of VaR computation. For the trading simulation task, the NNR model outperforms other models for the EUR/USD volatility, while the AR model based on squared returns performs best for the USD/JPY volatility. Finally, model combination and the inclusion of market data for currency volatility in 'mixed' models improve forecasting accuracy and VaR efficiency in most cases.

The rest of this paper is organised as follows. The second section presents a brief review of some previous literature relevant to this research. The third section describes the data, giving their statistical features. The fourth section depicts the benchmark models and alternative models estimated — giving a precise definition of both the time series models and the 'mixed' models investigated. The fifth section presents the estimation results for all the volatility models, focusing on out-of-sample results not only in terms of forecasting accuracy, but also in terms of risk management efficiency under the VaR framework and trading performance with a volatility filter strategy. The final section closes this paper with a summary of the conclusions.

LITERATURE REVIEW

Not surprisingly, as volatility is a key variable in asset pricing, asset allocation and financial risk management, there is a vast literature on volatility modelling, so just a brief review of recent articles relating to this research is given, mentioning also a few recent papers on the two advanced methods that are used in this study: NNR and SV modelling.

There is a wealth of articles supporting the use of generalised autoregressive conditional heteroscedasticity (GARCH) modelling for volatility forecasting (see, among others, Akgiray,[9] Bollerslev,[10] Bollerslev *et al.*,[11] Nelson,[12] Pagan and Schwert,[13] West and Cho[14]) and for financial applications such as VaR calculation (see, for instance, Andersen *et al.*,[15] Giot and Laurent[16]). Also, note that the popular RiskMetrics method, which was developed by JP Morgan[17] to compute VaR for risk management purposes, is derived from a standard GARCH(1,1) model. Among others, however, Neely and Weller[18] argue in favour of the use of genetic programming as an alternative technique to GARCH or RiskMetrics; Dunis and Huang[6] show that NNR and recurrent neural regression (RNN) models can significantly outperform GARCH models; while Wong *et al.*[19] also contend that GARCH-type models are not good enough for volatility forecasting and thus for managing market risk.

The SV model was originally suggested by Taylor,[20] Melino and Turnbull,[21] Harvey,[22] Harvey *et al.*[23] and Hamilton.[24] Recently, Yu[25] showed the superiority of the SV model for daily volatility forecasts of the NZSE40 index. Nonetheless, SV models are not yet as popular as GARCH-type models in empirical discrete-time finance applications. Bluhm and Yu[26] argue that the SV model is superior to GARCH-type models and other simpler models for option pricing but not for VaR applications and stock market trading strategies. Furthermore, Dunis *et al.*[27] and Dunis and Francis[28] indicate that the SV model underperforms more traditional approaches when forecasting FX volatility and the volatility of 10-year Government bonds.

Alternatively, Andersen *et al.*[29] contend that a parsimonious model can perform well in volatility forecasts in the presence of serial correlation in the standardised residuals or the squared standardised residuals. Mohammed[30] supports the use of autoregressive moving average (ARMA)(1,1) for FX volatility forecasts, an opinion shared by Pong *et al.*[31] Besides, Brooks and Persand[32] provide some evidence in favour of the AR model and the historical volatility model in the context of VaR estimation.

A growing literature has investigated whether non-linear effects are important in the conditional variance function. Fernandes[33] and Maheu and McCurdy[34] note that the realised FX volatility has non-linear features and thus GARCH(1,1) and SV models have only a limited predictive power for FX volatility owing to their linear specification. Dash and Kajiji[35] contend that the predictive power of a non-parametric NNR model is superior to that of GARCH with high frequency FX data. Moreover, Dunis and Huang[6] show that NNR and RNN models are the best single models in terms of FX volatility forecasting accuracy and in terms of option trading efficiency (for a broader discussion on NNR volatility applications, see also Azoff,[36] Bolland *et al.*,[37] Gradojevic and Yang,[38] Hu *et al.*,[39] Leung *et al.*,[40] Refenes,[41] Toulson,[42] White[43] and Zhang *et al.*[44]).

Admittedly, many researchers in finance have now come to the conclusion 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. For a while now, survey literature on forecast combinations, such as Clemen[45] and Mahmond,[46] has confirmed that combining different models generally provides more precise forecasts. Empirical results — such as in Granger and Ramanathan,[7] Russell and Adam,[47] Norman and Zeng[48] and Dunis *et al.*[27] — showed that combination methods do add value for forecasting accuracy. Moreover, Chapados and Bengio[49] document the advantages of using model combination for a VaR-based asset allocation using NNR models.

Finally, an important body of literature has investigated whether implied volatilities from the options markets are an unbiased and efficient predictor of ex-post realised volatility in both foreign exchange and equity markets (see, among others, Blair *et al.*,[50] Chiras and Manaster,[51] Christensen and Prabhala,[52] Giot,[53] Jorion[54] and Latane and Rendleman[55]). Canina and Figlewski[56] and Neely[57] provide opposite evidence against the use of implied volatilities alone. Bluhm and Yu[26] contend, however, that implied volatilities combined with GARCH models are a biased, but good predictor of German stock market volatility in the context of a trading strategy performance. Meanwhile, Dunis *et al.*[27] argue that 'mixed' models incorporating market data for currency volatility perform best most of the time in a medium-term FX volatility forecasting exercise, while Giot[53] further emphasises that a stock option implied volatility index adds value in terms of volatility forecasting and the implementation of a VaR model. This justifies the authors' willingness to complement the estimation of 'pure' time series models with implied volatility data obtained from the FX options market, leading to the estimation of 'mixed' time series models in order to test the null hypothesis that implied volatility does add value for volatility forecasting accuracy and risk management.

FX VOLATILITY AND RELATED FINANCIAL DATA

FX series and related financial data

The exchange rate series EUR/USD[58] and USD/JPY were extracted from a historical exchange rate database provided by Datastream. Logarithmic returns, defined as $S_t = \log(P_t/P_{t-1})$ were calculated for each exchange rate on a daily frequency. In order to approximate percentage changes, these logarithmic returns are multiplied by 100: $S_t = \log(P_t/P_{t-1}) \times 100$.

Moreover, it is important to consider the respective time zones and their implications for forecasting. For example, at the close of day *t* in European markets, eg at 17:00 hours, the closing data for day *t* in the American markets is unavailable as, in the US, markets would not close before 22:00 hours European time. Therefore it is necessary to introduce appropriate lags to reflect the time zone differences. Data that are not available can obviously not be used as a basis for forecasting.

All data were selected as possible explanatory variables to aid in the forecasting of the FX volatilities. A complete list of the data selected and the Datastream mnemonics is presented in Table 1.

Corresponding to the range of the implied volatility databank, the databank covers a span of more than five years, from 27th November, 1997 to 28th March, 2003. Because of lags, model estimation and validation are restricted to the period from 2nd January, 1998 to 28th March, 2003, leaving 1,339 trading days of observations for each exchange rate. The databank is further divided into two separate sets, with the first one covering 2nd January, 1998 to 13th May, 2002, for in-sample model estimation, and the second one from 14th May, 2002 to 28th March, 2003 (about one-sixth of the dataset, 223 datapoints) for out-of-sample model validation. Summary statistics for the

Table 1: Data and Datastream mnemonics

Number	*Variable*	*Mnemonic*
1	FTSE 100 — PRICE INDEX	FTSE100
2	DAX 30 PERFORMANCE — PRICE INDEX	DAXINDX
3	S&P 500 COMPOSITE — PRICE INDEX	S&PCOMP
4	NIKKEI 225 STOCK AVERAGE — PRICE INDEX	JAPDOWA
5	FRANCE CAC 40 — PRICE INDEX	FRCAC40
6	MILAN MIB 30 — PRICE INDEX	ITMIB30
7	DJ EUR STOXX 50 — PRICE INDEX	DJES50I
8	US EUR–$ 3-MONTH (LDN:FT) — MIDDLE RATE	ECUS$3M
9	JAPAN EUR-$ 3-MONTH (LDN:FT) — MIDDLE RATE	ECJAP3M
10	EUR EUR–CURRENCY 3-MONTH (LDN:FT) — MIDDLE RATE	ECEUR3M
11	GERMANY EUR–MARK 3-MONTH (LDN:FT) — MIDDLE RATE	ECWGM3M
12	FRANCE EUR–FRANC 3-MONTH (LDN:FT) — MIDDLE RATE	ECFFR3M
13	UK EUR–£ 3-MONTH (LDN:FT) — MIDDLE RATE	ECUK£3M
14	ITALY EUR–LIRE MONTH (LDN:FT) — MIDDLE RATE	ECITL3M
15	JAPAN BENCHMARK BOND–RYLD.10 YR (DS) — RED.YIELD	JPBRYLD
16	ECU BENCHMARK BOND–RYLD.10 YR (DS) — RED.YIELD	ECBRYLD
17	GERMANY BENCHMARK BOND 10 YR (DS) — RED.YIELD	BDBRYLD
18	FRANCE BENCHMARK BOND 10 YR (DS) — RED.YIELD	FRBRYLD
19	UK BENCHMARK BOND 10 YR (DS) — RED.YIELD	UKMBRYD
20	US TREAS.BENCHMARK BOND 10 YR (DS) — RED.YIELD	USBD10Y
21	ITALY BENCHMARK BOND 10 YR (DS) — RED.YIELD	ITBRYLD
22	Brent Crude — Current Month, fob US$/BBL	OILBREN
23	GOLD BULLION $/TROY OUNCE	GOLDBLN
24	Bridge/CRB Commodity Futures Index — PRICE INDEX	NYFECRB
25	US$ EUR (WMR) — EXCHANGE RATE	USEURSP
26	JAPANESE YEN TO US$ (WMR) — EXCHANGE RATE	JAPAYE$
27	US$ TO UK£ (WMR) — EXCHANGE RATE	USDOLLR
28	US$ TO Australia$ (WMR) — EXCHANGE RATE	AUSTDO$
29	Canadian$ TO US$ (WMR) — EXCHANGE RATE	CNDOLL$
30	HongKong$ TO US$ (WMR) — EXCHANGE RATE	HKDOLL$
31	Singapore$ TO US$ (WMR) — EXCHANGE RATE	SINGDO$
32	Chinese Yuan TO US$ (WMR) — EXCHANGE RATE	CHIYUA$

EUR/USD and USD/JPY daily returns, historical volatility and implied volatility series over the whole data period are presented in Table 2.

In line with the findings of many earlier studies on exchange rate changes (see,

Table 2: Summary statistics (2nd January, 1998–28th March, 2003)

	Log-returns		*Historical volatility*		*Implied volatility*	
	EUR/USD	*USD/JPY*	*EUR/USD*	*USD/JPY*	*EUR/USD*	*USD/JPY*
Mean	–0.000760	–0.002571	3.210382	3.840578	10.64310	12.19634
Median	–0.019634	0.002050	3.129101	3.539860	10.35000	11.20000
Maximum	1.442273	1.549575	5.759310	10.26845	16.75000	35.00000
Minimum	–0.984090	–2.858422	1.577985	2.115021	6.750000	7.300000
Std. dev.	0.266163	0.335824	0.788256	1.461081	1.959129	3.271973
Skewness	0.363250	–0.812038	0.487648	2.143690	0.593886	1.596657
Kurtosis	4.297349	9.027140	2.814678	8.335364	2.725505	7.287618
Jarque–Bera	123.3507	2173.868	54.98522	2613.713	82.91473	1594.577
Probability	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

among others, Baillie and Bollerslev,[59] Engle and Bollerslev,[60] West and Cho[14]), the statistics clearly show that FX logarithmic returns are non-normally distributed and heavily fat-tailed. So are the historical volatility and implied volatility. More specifically, as illustrated in Table 2, both FX returns have unconditional means not significantly different from zero, thus one can use squared returns as a measure of their variance and absolute returns as a measure of their standard deviation, which is common among market practitioners. Moreover, as suggested by Schwert,[61,62] the variance of a zero mean normally distributed variable is $\pi/2$ times the square of the expected value of its absolute value. Since the variables considered are not normally distributed, one can hence set this constant arbitrarily to 1. Taking, as a usual practice, a 252-trading day year, the one-month, ie 21-trading days, historical volatility is computed as the moving annualised standard deviation of returns

$$\mathrm{HVOL}_{21} = (1/21) \sum_{t-20}^{t} (\sqrt{252}) \times |S_t| \qquad (1)$$

where S_t stands for the EUR/USD or USD/JPY log-return at time t and HVOL_{21t} is the realised exchange rate volatilities over 21 days at time t, which the authors are interested in forecasting as accurately as possible.

Implied volatility series databank

Most studies dealing with implied volatilities, from Latane and Rendleman[55] to Giot,[53] have used data from listed options on exchanges rather than over-the-counter (OTC) volatility data. The problem in using exchange data is that call and put prices are only available for given strike levels. The corresponding implied volatility series must therefore be backed out using a specific option pricing model. As underlined by Dunis *et al.*,[27] this

Table 3: List of models for volatility forecasts

Model	*Description*	*Mnemonic*
1	Historical volatility	HistoricalVol
2	Implied volatility	ImVol
3	RiskMetrics volatility	RMVol
4	GARCH(p,q)	GARCH_type
5	GARCH(p,q) + implied volatility	GARCH_type
6	AR(p) based on absolute returns	AR_abs
7	AR(p) based on absolute returns + implied volatility	AR_abs
8	AR(p) based on squared returns	AR_sq
9	AR(p) based on squared returns + implied volatility	AR_sq
10	SV(1) based on log of squared returns	SV
11	SV(1) based on log of squared returns + implied volatility	SV
12	Neural Network + implied volatility	NNR
13	Average of all simple models except 'worst' models in-sample	Comb_avg
14	Average of all 'mixed' models except 'worst' models in-sample	Comb_avg
15	Regression-weighted average except 'worst' models in-sample	Comb_GR
16	Regression-weighted average of all 'mixed' models except 'worst' models in-sample	Comb_GR

procedure generates potential biases, such as material errors or mismatches. This is the reason why, as volatility is now an observable and traded quantity in many financial markets, this paper uses 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 no longer in option prices terms.[63]

The implied volatility time series used for the EUR/USD and USD/JPY were extracted from a market quoted implied volatility database originally provided by Chemical Bank (until the end of 1996) and updated from Reuters 'Ric' codes and subsequently maintained by CIBEF. These one-month at-the-money forward, market quoted volatilities are obtained from brokers by Reuters on a daily basis, at the close of business in London. Summary statistics for these implied volatility series over this restricted sample are shown in Table 2.

Admittedly, as discussed in Dunis *et al.*[27] and confirmed in Table 2, it is interesting to note that, on average, the mean level of implied volatilities is more than seven percentage points above that of historical volatility for both the EUR/USD and USD/JPY. It could, however, still be worth using implied volatility data directly available from the marketplace in order to improve forecasting accuracy: true, actual and implied volatility tend to move fairly closely together, which is indicated by their high correlation coefficient (over the period, the instantaneous correlation coefficient between historical and implied volatility is strongly positive and equal to 0.677 and 0.811 for EUR/USD and USD/JPY, respectively).

Further tests of autocorrelation, non-stationarity and heteroscedasticity (not reported here in order to conserve space) show that the FX log returns, historical volatility and implied volatility series are autocorrelated (except for the log returns), stationary and heteroscedastic over the whole observation period.

VOLATILITY FORECASTING MODELS

Table 3 gives a list of the 16 different models, both linear and non-linear, used for each time horizon considered. Each estimated time series model is complemented by a 'mixed' version counterpart integrating the additional information provided by implied volatility data. (The detailed specifications retained for each model and in-sample results are not reported here in order to conserve space; they are available from the authors upon request).

Benchmark models

Two naïve random walk models

Among the three benchmark models used in this paper, two simple naïve random walk models are first retained. One simply states that the best n-step ahead forecast of the conditional variance is its current past n-day average, while another one sets the n-step ahead forecast of the conditional variance at the current one-month implied volatility level. Consequently, the first type of 'naïve' model based on historical volatility yields the following n-step ahead forecast:

$$h_{t+n} = (1/\sqrt{252})\mathrm{HVOL}_{i,t} \tag{2}$$

where $\mathrm{HVOL}_{i,t}$ is the realised daily one-month historical volatility defined in equation (1).

The second type of 'naïve' model is based on market-quoted implied volatility and yields the following n-step ahead forecast:

$$h_{t+n} = (1/\sqrt{252})\mathrm{IMP}_{i,t} \tag{3}$$

where $\mathrm{IMP}_{i,t}$ is the implied daily one-month ($i = 21$) volatility prevailing at time t.

RiskMetrics volatility

The RiskMetrics volatility model[17] is also treated as a benchmark model owing to its popularity in risk measurement. Roughly speaking, RiskMetrics is one of the simplest tools for measuring financial market risk under the VaR framework. Derived from the GARCH(1,1) model, but with fixed coefficients, the RiskMetrics volatility is calculated using the standard formula

$$\mathrm{RMVOL}^2_{i,t} = \sigma^2_{(t/t-1)} = b\sigma^2_{(t-1)} + (1-b)\,.\,S^2_{(t)} \tag{4}$$

where σ^2 is the FX variance, $S^2_{(t)}$ is the FX squared return and $b = 0.94$ for daily data. This paper uses RiskMetrics volatility to forecast one-day, five-day and

21-day ahead for the out-of-sample period. The RiskMetrics volatility is calculated from equation (4), and then equation (5) is used to calculate the n-step ahead forecast

$$h_{t+n} = (1/\sqrt{252})\mathrm{RMVOL}_{i,t} \tag{5}$$

GARCH time series and 'mixed' models

The autoregressive conditional heteroscedasticity (ARCH) model was originally introduced by Engle[3] as a convenient way of modelling time-dependent conditional variance. It was later generalised by Bollerslev[10] and Taylor[20] as the GARCH model. Glossten *et al.*[64] and Zakoian[65] introduced TARCH (threshold ARCH) models as a simple extension of GARCH under the assumption that financial markets have an asymmetric response to news. Since the introduction of GARCH and its various extensions, hundreds of research papers have applied this modelling technique to measuring the volatility of financial time series. Basically, GARCH(1,1) 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 (the ARCH term) and the previous period's variance (the GARCH term). In the notation that has become standard, the GARCH(p,q) and TARCH(p,q) models are defined as:

$$h_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_j h_{t-j}^2 \tag{6}$$

$$h_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_j h_{t-j}^2 + \delta\varepsilon_{t-1} d_{t-1} \tag{7}$$

where $d_t = 1$ for $\varepsilon_t < 0$, and $d_t = 0$ otherwise, thus good news at time t has an impact of α_i, while bad news has an impact of $(\alpha_i + \delta)$.

Furthermore, the 'mixed' version counterparts of the GARCH(p,q) and TARCH(p,q), integrating implied volatility (IMP_t), yield the following formulation for the conditional variance (see, for instance, Kroner *et al.*[66]):

$$h_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_j h_{t-j}^2 + \gamma \mathrm{IMP}_{t-1} \tag{8}$$

$$h_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_j h_{t-j}^2 + \delta\varepsilon_{t-1} d_t + \gamma \mathrm{IMP}_{t-1} \tag{9}$$

Alternative GARCH-type models were tried for in-sample fitting in this research. Based on the Akaike Information Criterion/Schwarz Bayesian Criterion (AIC/SBC) information criteria, log-likelihood and standard error of the estimation, we ended up with a GARCH(1,1) and TARCH(1,1) for EUR/USD volatility simple and 'mixed' models respectively, while a GARCH(3,2) and GARCH(1,2) proved best for USD/JPY volatility simple and 'mixed' models, respectively.

Equations (6)–(9) give the one-step ahead forecast. For the TARCH n-step ahead forecast, as future values of d are unknown, $d = 0.5$ is arbitrarily set on the assumption that the distribution of the residuals is symmetric. The five-step and 21-step ahead forecasts can easily be obtained by recursive substitution. For instance, for GARCH(1,1),

$$\mathrm{E}_{(t)}[h_{t+n}^2] = \omega \sum_{i=1}^{n} (\alpha_1 + \beta_1)^{i-1} + [(\alpha_1 + \beta_1)^{n-1} h_t^2] \tag{10}$$

where $n = \{5, 21\}$ and h_t is obtained from equation (8).

AR(p) time series and 'mixed' models

As mentioned above, EUR/USD and USD/JPY squared returns could be regarded as a measure of their variance and absolute returns as a measure of their standard deviation. Furthermore, the stationarity of FX returns allow a traditional ARMA estimation procedure to be applied to the absolute and squared FX return series, provided that the presence of both heteroscedasticity and autocorrelation are allowed for where appropriate.

Because any moving average (MA) process can be represented in autoregressive form as an infinite AR process, but also because, in practice, out-of-sample n-step ahead forecasting with MA terms is not tractable, the process must be restricted to AR(p) processes. Actually, standard non-linear least squares estimation can be exploited, to correct for heteroscedasticity with the heteroscedasticity consistent covariance estimation proposed by White.[67] Following West and Cho,[14] the conditional variance based on absolute and squared returns is thus modelled as shown in equations (11) and (12).

$$h_t = \omega + \sum_{i=1}^{21} \alpha_i |S_{t-i}| \tag{11}$$

$$h_t = \omega + \sum_{i=1}^{21} \alpha_i S_{t-i}^2 \tag{12}$$

Based on the AIC/SBC information criteria, log-likelihood and standard error of the estimation, and taking into account only significant lags, we finally select restricted AR(4,5,9) and AR(1,6,9,18,20) processes, respectively for the conditional variance forecast based on squared returns for the EUR/USD and USD/JPY. For the conditional variance forecast based on absolute returns, the authors choose, respectively, restricted AR(4,5,9,14,17) and AR(1,2,4,7,9,13,14,19,20) processes. These lags represent a period of up to a maximum of four trading weeks.

The n-step ahead forecasts for AR(p) models are, respectively,

$$h_{t+n} = \omega + \sum_{i=1}^{21} \alpha_i |S_{t+n-i}| \tag{13}$$

$$h_{t+n} = \omega + \sum_{i=1}^{21} \alpha_i S_{t+n-i}^2 \tag{14}$$

The 'mixed' version counterpart of the AR(p) model with exogenous implied volatility variables is

$$h_t = \omega + \sum_{i=1}^{21} \alpha_i |S_{t-i}| + \gamma \mathrm{IMP}_{t-1} \tag{15}$$

$$h_t = \omega + \sum_{i=1}^{21} \alpha_i S_{t-i}^2 + \mathrm{IMP}_{t-1} \tag{16}$$

In terms of 'mixed' models, restricted AR(1,5) and AR(1,4,9,13,19) processes give the best in-sample results for the EUR/USD and USD/JPY volatility based on absolute returns, respectively, while we select AR(6) and AR(1,6,9,18,20) processes for the forecasts based on absolute returns. Again, the most updated information on implied volatility available at the time of the forecast $t-1$, IMP_{t-1} is applied to the n-step ahead forecast. Hence, the n-step ahead forecasts for the 'mixed' AR(p) models are, respectively

$$h_{t+n} = \omega + \sum_{i=1}^{21} \alpha_i |S_{t+n-i}| + \gamma \mathrm{IMP}_{t-1} \tag{17}$$

and

$$h_{t+n} = \omega + \sum_{i=1}^{21} \alpha_i S^2_{t+n-i} + \gamma \mathrm{IMP}_{t-1} \tag{18}$$

Stochastic variance SV(1) time series and 'mixed' models

FX volatility is time-dependent and model parameters are more likely to change over time than to stay constant, which could constitute a fatal drawback for fixed-parameter models. Intuitively, there is a clear attraction to the idea that volatility and its time-varying nature could be stochastic rather than the result of some deterministic function. Accordingly, this approach has recently drawn considerable attention, with a growing number of applications in finance, at least in academic circles if not among market practitioners.[68–71]

Broadly speaking, the SV model assumes that the variance is an unobservable process and volatility at time t, given all the information up to $t-p$, is random. Stochastic parameter regressions are based on a system of equations where the coefficient dynamics are usually modelled in a second equation. Thus it is possible to model volatility in state space form as a time-varying parameter model.

After several attempts at alternative specifications, the preferred approach was selected according to the log-likelihood, AIC criterion and the standard error of the estimation in-sample. Eventually, the authors chose to model the logarithm of the conditional variance as a random walk plus noise (note that working in logarithms ensures that h_t is always positive). They further assumed that the random coefficient, or 'state' variable, was best modelled as an AR(1) process with a constant mean, implying that shocks would show some persistence, but that the random coefficient would eventually return to its mean level, which is compatible with the behaviour of FX volatility.

$$\begin{aligned} \log(h_t) &= \omega + \mathrm{SV}_t + \varepsilon_t \\ \mathrm{SV}_t &= \delta \mathrm{SV}_{t-1} + \eta_t \end{aligned} \tag{19}$$

where SV is the time-varying coefficient, while ε_t and η_t are uncorrelated error terms.

Straightforwardly, the 'mixed' version counterpart of the SV(1) model with exogenous implied volatility variables yields

$$\begin{aligned} \log h_t &= \omega + \mathrm{SV}_t + \gamma \log \mathrm{IMP}_{t-1} + \varepsilon_t \\ \mathrm{SV}_t &= \delta \mathrm{SV}_{t-1} + \eta_t \end{aligned} \tag{20}$$

In order to derive the n-step ahead forecast for system (19), one must compute $\mathrm{E}(\mathrm{SV}_{t+n} | \mathrm{I}_t)$ with the information available at time t; it is clear from (19) that we have

$$\mathrm{E}(\mathrm{SV}_{t+1} | \mathrm{I}_t) = \delta \mathrm{E}(\mathrm{SV}_t) = \delta^2 \mathrm{SV}_{t-1} + \eta_t \tag{21}$$

Thus one can compute $\mathrm{E}(\mathrm{SV}_{t+n} | \mathrm{I}_t)$ by iterating equation (21)

$$\mathrm{E}(\mathrm{SV}_{t+n} | \mathrm{I}_t) = \delta^n \mathrm{SV}_t + \eta_t \tag{22}$$

One can now compute the n-step ahead forecast for the simple SV model as

$$\begin{aligned} \log h_{t+n} &= \omega + \mathrm{SV}_{t+n} + \varepsilon_{t+n} \\ &= \delta^n \mathrm{SV}_t + \eta_{t+n} \end{aligned} \tag{23}$$

Similarly, taking into account the fact that, in order to compute a truly out-of-sample forecast, the last information on implied volatility available at time $t-1$ is IMP_{t-1}, the 'mixed' system n-step ahead forecast becomes

$$\log(h_{t+n}) = \omega + \text{E}(\text{SV}_{t+n} \mid \text{I}_t) + \gamma\log(\text{IMP}_{t-1}) + \varepsilon_t \quad (24)$$
$$\text{E}(\text{SV}_{t+n} \mid \text{I}_t) = \delta^n \text{SV}_t + \eta_t$$

NNR models

NNR regression models, in particular, have been applied with increasing success to economic and financial forecasting and would, according to some, constitute the state of the art in forecasting methods (see, for instance, Zhang *et al.*[44]).

It is well beyond the scope of this paper to give a complete overview of NNR models, their biological foundation and their many architectures and potential applications. This paper uses exclusively the multilayer perceptron, a multilayer feedforward network trained by error back propagation. For a full discussion of NNR models, refer to Haykin,[72] Kaastra and Boyd,[73] Kingdon[74] and Zhang *et al.*[44].

For present purposes, suffice it 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.

Theoretically, the advantage of neural networks over traditional forecasting methods is that, as is often the case, the model best adapted to a particular problem cannot be identified beforehand. It is then better to resort to a method that is a generalisation of many models, than to rely on an *a priori* model.

Successful applications in forecasting foreign exchange rates can be found in Deboeck,[75] Kuan and Liu[76] and Franses and Van Homelen[77] among others, while Dunis and Huang[6] show the benefits of NNR and RNN models in terms of FX volatility forecasting accuracy and in terms of option trading efficiency.

Developing NNR models is a rather difficult and time-consuming task. For this research, it was necessary to develop one NNR model per forecast horizon (one-day, five-days or 21-days ahead) with different appropriate lagged input variables for both the EUR/USD and USD/JPY volatilities. In the circumstances, only the 'mixed' models were estimated to check whether NNR models outperform the other 'mixed' models or not. The detailed procedure followed is documented in Dunis and Huang[6].

Following standard heuristics to reduce the risk of overfitting and to control the error, the total data set was divided into approximately two-thirds for training, one-sixth for the test period and one-sixth for the validation set, which, as for the other models, is the out-of-sample period dataset. Both the training and the following test period are used for model tuning: the training set is used to develop the model, while the test set measures how well the

Table 4: Explanatory variables of one-day ahead NNR forecasting model

EUR/USD vol. *Variable*	*Best lags*	*USD/JPY vol.* *Variable*	*Best lags*
RMVol_EURUSD	1	IMV30USDJPY	2
IMV30EURUSD	9	ABS_USDJPY	10
ABS_EURUSD	20	ABS_USDJPY	20
ABS_EURUSD	1	RMVol_USDJPY	21
ECUS$3M	11	AUSTDO$	2
DAXINDX	20	ECITL3M	9
DJES50I	20	BDBRYLD	21
FRCAC40	20	ECBRYLD	21
ITMIB30	19	S&PCOMP01	3
NYFECRB	10	GOLDBLN	3
JPBRYLD	10		
SINGDO$	10		
ECITL3M	20		

model interpolates over the training set and makes it possible to check during the adjustment whether the model remains valid for the future. The validation set is used to estimate the actual performance of the model in a deployed environment.

Inputs are transformed into returns: despite some contrary opinions, eg Balkin,[78] stationarity remains important if NNR models are to be assessed on the basis of the level of explained variance (see Dunis and Huang[6]). In the absence of an indisputable theory of FX volatility, it is assumed that it could be explained by that FX recent evolution, lagged RiskMetrics volatility, volatility spillovers from other financial markets and the yield curve (computed as the difference between ten-year bond yields and three-month interest rates) as a measurement of macroeconomic and monetary policy expectations. Final inputs and lags of the one-day ahead NNR volatility forecasting model are presented in Table 4. The one-day ahead NNR model for the EUR/USD uses 13 inputs and one hidden layer with six nodes, while the NNR model for the USD/JPY uses ten inputs and one hidden layer with five nodes.

All variables are normalised according to the choice of the sigmoid activation function. Commencing from a traditional linear correlation analysis, variable selection is achieved via a forward stepwise neural regression procedure. Starting with lagged implied volatility, other potential input variables are progressively added, keeping the network architecture constant. If adding a new variable improves the level of explained volatility over the previous 'best' model, the pool of explanatory variables is updated. If there is a failure to improve over the previous 'best' model after several attempts, variables in that model are alternated to check whether no better solution can be achieved. The chosen model is then kept for further tests and improvements.

Combined time series and 'mixed' models

Many researchers in finance have now come to the conclusion 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.

Survey literature on forecast combinations such as Clemen[45] and Mahmond[46] have confirmed that combining different models generally provides more precise forecasts. This statement on the advantages of combining two or more forecasts into a composite forecast is consistent with findings by Makridakis *et al.*,[79] Granger and Ramanathan,[7] Diebold and Lopez,[80] Dash and Kajiji[35] and Dunis and Huang,[6] among others. These articles agree that model combination of several methods improves overall forecasting accuracy over and above that of the individual forecasting models used in the combination. Accordingly, there is a strong case for combining the various models retained in this research, and this is why two different, yet simple, model combinations are computed.

Simple average model combination

The first forecast combination retained is the simple average of each single forecasting model for time $t + n$, minus the model which performs worst when the in-sample forecasting accuracy measures at the one-day horizon are analysed: the naïve implied volatility when considering the 'pure' time series model for both FX volatilities and the 'mixed' SV model for both FX volatility forecasts when considering 'mixed' time series models. Thus:

$$h_{t+n} = (1/m) \sum_{i=1}^{m} h_{i,t+n} \quad (25)$$

where $n = \{1, 5, 21\}$ and $h_{i,t+n}$ represents the predicted volatility of the m single forecasting models for time $t + n$.

Regression-weighted average combination

The second forecast combination uses the linear regression weighting approach suggested by Granger and Ramanathan (again excluding the worst model in-sample),[7] which yields

$$h_{t+n} = a + \sum_{i=1}^{m} b_i h_{i,t+n} \quad (26)$$

where $n = \{1, 5, 21\}$ and $h_{i,t+n}$ represents the predicted volatility of the m single forecasting models for time $t + n$.

Based on the one-day horizon in-sample results, only models with significant weights are retained. For the 'pure' time series models, two models are selected for the EUR/USD volatility forecasts, ie the AR(p) model based on absolute returns and the historical volatility, and four models for the USD/JPY volatility forecasts, ie the two AR(p)-based models, the SV model and historical volatility. For the 'mixed' models including implied volatility, two models for the EUR/USD volatility forecasts, ie the AR(p) model based on squared returns and the NNR model, and three models for the USD/JPY volatility forecasts, ie the two AR(p)-based models and the NNR model were retained.

OUT-OF-SAMPLE ESTIMATION RESULTS

Comparison criteria

Market participants who can reliably predict volatility should be able to profit from their superior forecasting ability from a trading perspective and to control better the financial risks associated with their positions in terms of risk management within a VaR context.

Consequently, the model comparison criteria are not limited to statistical forecasting accuracy. The authors also focus on applying a simple VaR model using the out-of-sample volatility forecasts derived from the models and thus compare their ability to estimate VaR. Moreover, in order to provide a more direct assessment of the economic value of each model, a trading simulation using a volatility filter is implemented. As mentioned above, the out-of-sample period from 14th May, 2002 to 28th March, 2003, is used for model evaluation and comparison.

Forecasting accuracy measures

As is standard in the economic literature, the RMSE, the MAE, the MAPE and Theil-U statistic are computed. These measures have already been presented in details by, among others, Makridakis *et al.*,[81] Pindyck and Rubinfeld[82] and Theil.[83] Following Dunis and Huang,[6] we also compute a CDC measure, which checks whether the direction given by the forecast is the same as the actual change which subsequently occurred (ie the direction of change implied by the forecast at time t for time $t + n$ compared with the volatility level prevailing at time t).

The RMSE and MAE statistics are scale-dependent measures but give a basis to compare volatility forecasts with the realised volatility. The MAPE, Theil-U and CDC statistics are independent of the scale of the variables. In particular, the Theil-U statistic is constructed in such a way that it necessarily lies between zero and one, with zero indicating a perfect fit, whereas the CDC lies by construction between 0 and 100 per cent, the latter indicating a perfect forecast of changes (note that a CDC of 50 per cent is the random result and values below 50 per cent imply a worse than random performance).

For four of the error statistics retained (RMSE, MAE, MAPE and Theil-U), the lower the output, the better the forecasting accuracy of the model concerned. Rather than depending on securing the lowest statistical forecast error, however, the profitability of a trading system critically depends on taking the right position and therefore getting the direction of changes right. RMSE, MAE, MAPE and Theil-U are all important error measures, yet they may not constitute the best criteria from a profitability point of view. The CDC statistics address this issue and, for this measure, the higher the output the better the forecasting accuracy of the model concerned.

VaR efficiency

VaR is a standard quantitative tool for estimating, over a given period, the potential loss on a financial portfolio with a given probability level. The actual VaR measure is based on a volatility forecast, generally one period ahead as VaR is now widely used by financial institutions to calculate their overall market risk at the end of each trading day (see Basle Committee,[84,85] Hendricks,[86] Marshall and Siegel,[87] Hull and White,[88] Jorion,[89] Linsmeier and Pearson[90] and Janssen[91] for a full discussion).

In this application, different volatility forecasts are fed into the basic VaR model. At the two confidence levels of 1 per cent and 5 per cent, respectively, under the normality assumption, VaR is computed as follows:

$$\mathrm{VaR}_{i,t+n}^{a=1\%} = 2.326\hat{\sigma}_{i,t+n} \tag{27}$$

$$\mathrm{VaR}_{i,t+n}^{a=5\%} = 1.645\hat{\sigma}_{i,t+n} \tag{28}$$

where $\hat{\sigma}_{i,t+n} = h_{i,t+n}$ are the FX volatility forecasts derived from models 1-16 (see Table 3) in each case.

As FX traders can either buy or sell a currency, one needs to consider both tails of the return distribution (ie both positive and negative returns). One would therefore expect a 'hit rate' of 2 per cent and 10 per cent, respectively (as the 'hit rate' is the percentage occurrence of an actual loss greater than the predicted maximum loss in the VaR framework, assuming normal return distributions; the 'hit rates' based on two-tailed probability levels have expected values of 2 per cent and 10 per cent at the 1 per cent and 5 per cent confidence levels, respectively). Consequently, in this application, a volatility model with a 'hit rate' close to the expected one is regarded as a 'good' model in terms of VaR efficiency for risk management.[92] The 'hit rate' $H_{i,n}$ of model i at horizon n is given by:

$$H_{i,n} = 100 \times \frac{1}{N} \sum_{i=1}^{n} \begin{cases} 1 & \text{if} |S_{t+n}| > \mathrm{VaR}_{i,t+n} \\ 0 & \text{otherwise} \end{cases} \tag{29}$$

where N is the number of trading days out-of-sample.

Trading performance

As mentioned before, a superior volatility forecasting ability should translate into a superior options trading performance, an assertion analysed by Dunis and Huang,[6] who apply a realistic volatility trading strategy using FX option straddles once mispriced options have been identified.

The idea of the trading simulation hereafter is simpler: it is based on trading strategies using the one-day ahead volatility forecast and combining naïve, moving average convergence divergence (MACD) and exponential moving average (EMA) systems with volatility filters.[93] It is important to note that the simulation is only aimed at checking which forecasting models are efficient for risk management purposes through the application of a volatility filter strategy, instead of for optimising the performance of trading rules with the addition of volatility filters. Accordingly, the signals are defined as:

$$\text{Naïve signal as} \begin{cases} \text{long} & \text{if } S_{t-1} > 0 \\ \text{short} & \text{otherwise} \end{cases} \tag{30}$$

$$\text{MACD signal as} \begin{cases} \text{long} & \text{if } (1/n)\sum_{i=1}^{n} S_{t-n} > (1/m)\sum_{i=1}^{m} S_{t-n} \\ \text{short} & \text{otherwise} \end{cases} \tag{31}$$

where $n = 10$ and $m = 20$ (arbitrarily retained here).

$$\text{EMA signal as} \begin{cases} \text{long} & \text{if } \text{MA}_t > 0 \\ \text{short} & \text{otherwise} \end{cases} \tag{32}$$

where $\text{MA}_t = \text{MA}_{t-1} + \alpha\{S_t - \text{MA}_{t-1}\}$, with $\text{MA}_0 = \text{S}_0$ and $\alpha = 1/n$ (ie the inverse of the chosen time span); thus the higher α, the smaller the smoothing effect. The trading simulation arbitrarily retains $n = 8$ trading days, ie $\alpha = 0.125$.

It is well known that MACD and EMA systems perform poorly in volatile markets, precisely because volatile markets imply frequent direction changes. A volatility filter is therefore introduced which makes good for such volatile periods and overlays the signals given by Equations (30)–(32). In other words, instead of having systems that are constantly in the market, long or short, there are also periods where the systems have no position. The volatility signal is given in Equation (33), where $E(h_{i,t+1})$ is the predicted volatility from model i, $i = \{1,16\}$ for time $t + 1$.

$$\text{Volatility filter signal} \begin{cases} 1\ (= \text{model position}) & \text{if } E(h_{i,t+1}) < \text{volatility filter} \\ 0\ (= \text{no position}) & \text{otherwise} \end{cases} \tag{33}$$

Among the list of conventional trading performance measures used by the fund management industry to analyse trading results (see, among others, Dunis and Jalilov[94] and Dunis and Williams[95] for details), the present authors focus on the annualised return; the Sharpe ratio (a measure of risk-adjusted return); the maximum drawdown (a measure of downside risk showing the maximum cumulative loss that could have been incurred on a portfolio); the average gain/loss ratio; and the probability of a 10 per cent loss.

Finally, as this paper is more concerned with the task of singling out models that are efficient for risk management purposes through the application of a volatility filter strategy, rather than truly optimising the performance of trading rules, no transaction costs are considered in this simulation.

Model rankings

Choosing the best models is not such a simple matter, as the 'best' model is dependent upon the choice of criteria. In order to rank the models according to statistical forecasting accuracy, a score is given to each accuracy measure, a score of 1–9 for the nine simple models (1–7 in the case of the 'mixed' models) to each RMSE, MAE, MAPE and Theil-U, and a score half that size for CDC: the original weight of the CDC is halved as it is the only measure of direction, a key criterion in financial markets (see Dunis and Francis[28]). For example, the best model in terms of RMSE gets a score of 1, the second best a score of 2 and so on, while, for the CDC, the model with the highest CDC gets a score of 0.5, the second best a score of 1 and so on, so that in the end, the model with the lowest points total is chosen as the best one.

The same ranking approach is used for the VaR efficiency and the trading performance applications (note that, for the trading application, all six trading performance measures are equally weighted).

OUT-OF-SAMPLE ESTIMATION RESULTS

Having documented the different measures that are used to estimate the forecasting accuracy and application performance of the different models, the out-of-sample results are now analysed. The authors basically wish to answer the following questions:

(1) How do the models fit out-of-sample and is there a (or several) good forecasting model(s) in the context of both forecasting accuracy and the applications?
(2) Do model combinations and implied volatility data add value in terms of forecasting accuracy and in terms of the applications retained?

Again, in order to conserve space, only some of the results are reported (in the Appendix), but complete results are available from the authors upon request.

Forecasting accuracy results

The results of the volatility models are somewhat mixed. Most indeed have a certain forecasting power, as proved for instance by the Theil-U statistics, but errors often remain important, as some models register MAPE levels of over 100 per cent and less than 50 per cent CDC.

Beginning with the pure time series models, it is worth noting that implied volatility is the worst forecaster for both FX volatilities at most horizons (see Appendix A1, Table A1.1). Hence, all the other volatility forecasting models offer more precise indications about future volatility than implied volatilities. For the one-day ahead forecast, surprisingly, historical volatility outperforms all other models for the EUR/USD volatility. Historical volatility also comes first for USD/JPY volatility, followed by the regression-weighted combination. The five-day ahead forecasts show similar results. For the 21-day horizon, the regression-weighted combination outperforms the other methods for the EUR/USD volatility while, for the USD/JPY volatility, the RiskMetrics volatility is the best forecaster, followed by historical volatility.

Turning to the 'mixed' models, at the one-day horizon, the NNR model and the two model combinations are the best for the EUR/USD and USD/JPY volatility forecasts, respectively. Similar results can be found at the five- and 21-day horizons, with NNR and model combinations generally preferred (see Appendix A1, Table A1.2).

In summary, the results for pure time series models indicate that historical volatility is a parsimonious technique quite difficult to beat, while NNR models do add value when considering the 'mixed' models approach. Moreover, model combinations generally help to improve forecasting accuracy (see Appendix A1, Table A1.3). Not surprisingly, the forecasting error increases with longer horizons. Finally, the results also show that the USD/JPY volatility is comparatively more difficult to forecast.

VaR implementation results

Some models have a 'hit rate' far from the expected value. Besides, for models which tend to constantly overestimate volatility, the daily return may never or seldom reach the computed VaR threshold (ie they would have a zero 'hit rate', or close to that level). Conversely, for models constantly underestimating volatility, the daily return would continuously breach the VaR level, yielding a 'hit rate' of 100 per cent: such 'hit rates' obviously have little economic significance and are therefore excluded from the rankings.

Looking at both the one per cent and five per cent significance levels, for the pure time series models at a one-day horizon, the AR model based on absolute returns is the best model for computing the EUR/USD VaR, followed by the AR model based on squared returns and the simple average model combination. Interestingly, historical and RiskMetrics volatility are ranked as poor models. Similar results can be found for the USD/JPY VaR computation. At the five-day horizon, the results show the simple average model combination performing best of all. At the 21-day horizon, the simple average model combination remains the best VaR computation method, while the two benchmark models — historical and RiskMetrics volatility — perform quite poorly (see Appendix A2, Table A2.1).

Moving to the 'mixed' models, the result of the EUR/USD VaR computation is rather similar at the three different horizons, with the AR model based on squared returns the best model available, closely followed by the NNR model. It is worth noting that GARCH models perform quite poorly for all horizons. Turning to the USD/JPY VaR computation, the results are quite surprising, with the NNR model and the regression-weighted model combination almost always worse than others for all horizons, while the AR model based on squared returns is generally considered best (see Appendix A2, Table A2.2).

On the whole, in terms of preferred volatility models, the results of the VaR implementation do not confirm those of the forecasting accuracy tests. Roughly speaking, the simple average model combination comes first among the pure time series models while, for 'mixed' models including implied volatility, the AR model based on squared returns is best. As could be expected, the best 'mixed' model outperforms the best simple model (see Appendix A2, Table A2.3). Finally, the VaR application shows that model combination does add value.

Trading performance results

Models that constantly overestimate realised volatility, such as implied volatility, are not considered for the trading simulation incorporating a volatility filter. Looking at the trading performance results, one can see that some volatility models indeed manage to reduce risk and generate higher profits, as evidenced by the Sharpe ratio and average gain/loss ratio. Still, some combined models prove disappointing while, in some cases, the volatility filter prevents any transaction, as an overestimated volatility forecast well beyond the chosen filter level prevents the underlying model from trading throughout the out-of-sample period.

Starting with the trading performance derived from pure time series models for the EUR/USD, the empirical results show that the AR model based on absolute returns is the best one for the naïve and MACD strategies overlayed with a volatility filter while, for the EMA strategy, the regression-weighted average model combination outperforms other approaches. Looking at the USD/JPY, historical volatility, which was best in terms of forecasting accuracy, is generally inferior to all other filtering methods, while model combination, GARCH and RiskMetrics are the best filtering techniques for the naïve, MACD and EMA strategies, respectively (see Appendix A3, Table A3.1).

Moving to the performance derived from 'mixed' models adding implied volatility information as an extra explanatory variable, NNR models appear superior in most cases for the EUR/USD (with the regression-weighted average preferred for the EMA strategy — see Appendix A3, Table A3.2). Turning to the USD/JPY, the AR model based on squared returns outperforms other 'mixed' models in providing the best filter for the naïve and EMA trading strategies, while coming a close second to the simple average for the MACD strategy. Note that the NNR and SV models do not perform well in terms of volatility filter for the USD/JPY trading strategies.

On the whole, the results in Appendix A3 show that volatility filters do add value to the three basic trading strategies retained. Filters derived from 'mixed' volatility models are particularly useful for improving trading results, with the NNR and the AR model based on the squared returns model the best single modelling approaches for the EUR/USD and the USD/JPY, respectively, over the out-of-sample period.

CONCLUSION

This paper investigated the predictability of 16 alternative volatility models applied to the EUR/USD and USD/JPY exchange rates for risk management and trading purposes. The forecasting accuracy of the predictions of the models retained was analysed, but the main concern is whether these forecasts can improve financial risk management. Therefore, the financial applicability of these forecasts was estimated within a simple VaR framework and in a trading simulation, using a volatility filter strategy for risk management. Meanwhile, whether implied volatility information available from the marketplace adds value, and whether model combination can help improve forecasting accuracy and risk management were also investigated.

All 16 models were developed over the period January 1998 to mid-May 2002, to be applied over the same out-of-sample period from 14th May, 2002 to 28th March, 2003. It must therefore be stressed that, with 16 volatility models, 223 out-of-sample forecasts at three horizons for two currencies, a total of over 20,000 forecasts[96] were produced. In order to keep the project manageable, all models were selected on the basis of providing the best fit for both FX volatilities concerned over the in-sample dataset. The specification of the volatility models was then kept constant during the entire forecast period. Adjustment of the specification was therefore not allowed during the course of the forecasting exercise but, given the time-varying nature of volatility, it is probably safe to assume that respecifying the models during the forecast period would have led to an increase in forecasting accuracy.

In any case, the empirical results clearly show that statistical forecasting accuracy is not the only key to VaR efficiency or trading performance — something already pointed out in recent research such as Bluhm and Yu,[26] Dunis and Huang[6] and Wong *et al.*[19]

As it is, the results show that, if no single volatility model emerges as an overall winner in terms of forecasting accuracy, risk management efficiency and FX trading performance, 'mixed' models incorporating market data for currency volatility, NNR models and model combination perform best most of the time. As an example, the single NNR model does improve forecasting accuracy and gives the best results in terms of trading performance for the EUR/USD, even if its performance in the VaR application and for the USD/JPY trading simulation is somewhat disappointing.

Model combination generally improves forecasting accuracy and VaR efficiency, yet it does not seem to help much in the trading application, where single models seem to perform better overall. Still, more often than not, 'mixed' models incorporating implied volatility information from the marketplace appear as good performers in terms of forecasting accuracy, risk management and trading, in particular for the EUR/USD. This paper, therefore, rejects the null hypothesis that implied volatility does not add value in improving forecasting accuracy and risk management, which is consistent with the findings of Blair *et al.*,[50] Giot[53] and Dunis *et al.*[27] among others.

References and notes

1 Garman, M. B. and Kohlhagen, S. W. (1983) 'Foreign Currency Option Values', *Journal of International Money and Finance*, Vol. 2, pp. 231–237.

2 Brailsford, T. J. and Faff, R. W. (1996) 'An Evaluation of Volatility Forecasting Techniques', *Journal of Banking and Finance*, Vol. 20, pp. 419–438.

3 Engle, R. F. (1982) 'Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation', *Econometrica*, Vol. 50, pp. 987–1007.

4 Engle, R. F., Hong, C. H., Kane, A. and Noh, J. (1993) 'Arbitrage Valuation of Variance Forecasts with Simulated Options', in Chance, D. M. and Tripp, R. R. (eds), 'Advances in Futures and Options Research', JIA Press, Greenwich, CT.

5 Laws, J. and Gidman, A. (2000) 'Forecasting Stock Market Volatility and the Application of Volatility Trading Models', CIBEF Research Paper, Liverpool Business School, Liverpool, UK.

6 Dunis, C. and Huang, X. (2002) 'Forecasting and Trading Currency Volatility: An Application of Recurrent Neural Regression and Model Combination', *Journal of Forecasting*, Vol. 21, pp. 317–354.

7 Granger, C. W. J. and Ramanathan, R. (1984) 'Improved Methods of Combining Forecasts', *Journal of Forecasting*, Vol. 3, pp. 197–204.

8 BIS (2004) 'Triennial Central Bank Survey of Foreign Exchange and Derivatives Market

Activity in April 2004', September, BIS, Basle, Switzerland.
9 Akgiray, V. (1989) 'Conditional Heteroskedasticity in Time Series of Stock Returns: Evidence and Forecasts', *Journal of Business*, Vol. 62, pp. 55–80.
10 Bollerslev, T. (1986) 'Generalized Autoregressive Conditional Heteroskedasticity', *Journal of Econometrics*, Vol. 31, pp. 307–327.
11 Bollerslev, T., Chou, R. Y. and Kroner, K. F. (1992) 'ARCH Modelling in Finance: A Review of the Theory and Empirical Evidence', *Journal of Econometrics*, Vol. 52, pp. 5–59.
12 Nelson, D. (1991) 'Conditional Heteroskedasticity in Asset Returns: a New Approach', *Econometrica*, Vol. 59, pp. 349–370.
13 Pagan, A. and Schwert, G. (1990) 'Alternative Models for Conditional Stock Volatilities', *Journal of Econometrics*, Vol. 69, pp. 367–391.
14 West, K. D. and Cho, D. (1995) 'The Predictive Ability of Several Models of Exchange Rate Volatility', *Journal of Econometrics*, Vol. 69, pp. 367–391.
15 Andersen, G. T., Bollerslev, T., Diebold, F. X. and Labys, P. (2003) 'Modelling and Forecasting Realized Volatility', *Econometrica*, Vol. 71, No. 2, pp. 579–625.
16 Giot, P. and Laurent, S. (2001) 'Modelling Daily Value-at-Risk Using Realized and ARCH-Type Models', *Journal of Banking and Finance*, Vol. 26, pp. 1273–1296.
17 JP Morgan (1994) 'RiskMetrics Technical Document', Morgan Guaranty Trust Company, New York, NY.
18 Neely, C. J. and Weller, P. A. (2002) 'Predicting Exchange Rate Volatility: Genetic Programming Versus GARCH and RiskMetrics™', *Federal Reserve Bank of St. Louis Review*, Vol. 84, No. 3, pp. 43–54.
19 Wong, M., Cheng, W. Y. and Wong, C. (2003) 'Market Risk Management of Banks: Implications from the Accuracy of Value-at-Risk Forecasts', *Journal of Forecasting*, Vol. 22, pp. 23–33.
20 Taylor, S. (1986) 'Modelling Financial Time Series', John Wiley, Chichester, UK.
21 Melino, A. and Turnbull, S. M. (1990) 'Pricing Foreign Currency Options with Stochastic Volatility', *Journal of Econometrics*, Vol. 45, pp. 239–265.
22 Harvey, A. C. (1991) 'Forecasting, Structural Time Series Models and the Kalman Filter', Cambridge University Press, Cambridge, UK.
23 Harvey, A., Ruiz, E. and Shephard, N. (1994) 'Multivariate Stochastic Variance Models', *Review of Economic Studies*, Vol. 61, pp. 247–264.
24 Hamilton, J. D. (1994) 'Time Series Analysis', Princeton University Press, Princeton, NJ.
25 Yu, J. (2002) 'Forecasting Volatility in the New Zealand Stock Market', *Applied Financial Economics*, Vol. 12, No. 3, pp. 193–202.
26 Bluhm, H. H. W. and Yu, J. (2001) 'Forecasting Volatility: Evidence from the German Stock Market', International Conference on Modelling and Forecasting Financial Volatility, University of Western Australia, Perth, 7th–9th September.
27 Dunis, C., Laws, L. and Chauvin, S. (2003) 'FX Volatility Forecasts and the Informational Content of Market Data for Volatility', *European Journal of Finance*, Vol. 9, No. 3, pp. 242–272.
28 Dunis, C. and Francis, F. L. (2004) 'The Information Content of Swaption Rates for USD and EUR Government Bonds Volatility Models', *Derivatives Use, Trading & Regulation*, Vol. 10, No. 3, pp. 197–228.
29 Andersen, G. T., Bollerslev, T., Diebold, F. X. and Labys, P. (1999) 'The Distribution of Exchange Rate Volatility', *NBER Working Paper Series*, No. 6961, NBER, Washington, DC.
30 Mohammed, A. A. A. (1999) 'On Forecasting Exchange Rates: A Time Series Analysis', IAOS Conference Paper.
31 Pong, S., Shackleton, M. B., Taylor, S. J. and Xu, X. (2003) 'Forecasting Currency Volatility: A Comparison of Implied Volatilities and AR(FI)MA Models', *Journal of Banking and Finance*, Vol. 28, No. 10, pp. 2541–2564.
32 Brooks, C. and Persand, G. (2003) 'Volatility Forecasting for Risk Management', *Journal of Forecasting*, Vol. 22, pp. 1–22.
33 Fernandes, M. (1998) 'Non-linearity and Exchange Rates', *Journal of Forecasting*, Vol. 17, No. 7, pp. 497–514.
34 Maheu, J. M. and McCurdy, T. H. (2002) 'Nonlinear Feature of Realized FX Volatility', *Review of Economics and Statistics*, Vol. 84, No. 4, pp. 668–681.
35 Dash, G. H. and Kajiji, N. (2001) 'Prediction of FX Volatility via an RBF Neural Network with Closed-Form Regularization', 8th International Conference on Forecasting Financial Markets, May, London, UK.
36 Azoff, E. M. (1994) 'Neural Network Time Series Forecasting of Financial Markets', John Wiley, Chichester, UK.
37 Bolland, P. J., Connor, J. T. and Refenes, A.-P. N. (1998) 'Application of Neural Networks to Forecast High Frequency Data: Foreign Exchange', in Dunis, C. and Zhou, B. (eds), 'Nonlinear Modelling of High Frequency

Financial Time Series', John Wiley, Chichester, UK, pp. 225–246.
38 Gradojevic, N. and Yang, J. (2000) 'The Application of Artificial Neural Networks to Exchange Rate Forecasting: The Role of Market Microstructure Variables', Bank of Canada Working Paper, RePEc: bca:bocawp:0–23.
39 Hu, M. Y., Zhang, G., Jiang, C. X. and Patuwo, B. E. (1999) 'A Cross-Validation Analysis of Neural Network Out-of-Sample Performance in Exchange Rate Forecasting', *Decision Sciences*, Vol. 30, No. 1, pp. 197–216.
40 Leung, M. T., Daouk, H. and Chen, A. -S. (2000) 'Forecasting Stock Indices: A Comparison of Classification and Level Estimation Models', *International Journal of Forecasting*, Vol. 16, pp. 173–190.
41 Refenes, A. N. (1993) 'Constructive Learning and its Application to Exchange Rate Forecasting', in Trippi, R. and Turban, E. (eds), 'Neural Networks in Finance and Investing — Using Artificial Intelligence to Improve Real-Word Performance', Probus, Chicago, IL.
42 Toulson, S. (1996) 'Forecasting Level and Volatility of High Frequency Exchange Rate Data', *Chemical Bank Working Papers in Financial Economics*, Vol. 9.
43 White, H. (1988) 'Economic Prediction Using Neural Networks: The Case of IBM Daily Stock Returns', *Proc. IEEE International Conference on Neural Networks*, San Diego, Vol. 2, CA, pp. 451–459.
44 Zhang, G., Patuwo, B. E and Hu, M. Y. (1998) 'Forecasting with Artificial Neural Networks: The State of the Art', *International Journal of Forecasting*, Vol. 14, pp. 35–62.
45 Clemen, R. T. (1989) 'Combining Forecasts: A Review and Annotated Bibliography', *International Journal of Forecasting*, Vol. 5, pp. 559–583.
46 Mahmond, E. (1984) 'Accuracy in Forecasting', *Journal of Forecasting*, Vol. 3, pp. 139–160.
47 Russell, T. D. and Adam, Jr, E. E. (1987) 'An Empirical Evaluation of Alternative Forecast Combinations', *Management Science*, Vol. 33, pp. 1267–1276.
48 Norman, R. S. and Zeng, T. (2001) 'Choosing among Competing Econometric Forecasts: Regression-Based Forecast Combination Using Model Selection', *Journal of Forecasting*, Vol. 20, pp. 425–440.
49 Chapados, N. and Bengio, Y. (2002) 'Cost Functions and Model Combination for VaR-Based Asset Allocation Using Neural Networks', CIRANO Working Paper, 2002s–49 Université de Montréal, Montréal, Canada.
50 Blair, B. J., Poon, S. and Taylor, S. J. (2001) 'Forecasting S&P100 Volatility: The Incremental Information Content of Implied Volatilities and High-Frequency Index Returns, *Journal of Econometrics*, Vol. 105, pp. 5–26.
51 Chiras, D. P. and Manaster, S. (1978) 'The Information Content of Option Prices and a Test of Market Efficiency', *Journal of Financial Economics*, Vol. 6, pp. 213–234.
52 Christensen, B. J. and Prabhala, N. R. (1998) 'The Relation Between Implied and Realized Volatility', *Journal of Financial Economics*, Vol. 50, pp. 125–150.
53 Giot, P. (2002) 'The Information Content of Implied Volatility Indices for Forecasting Volatility and Market Risk', CORE Discussion Paper, 2003 (27), Center for Research in Finance and Management, Université Catholique de Louvain, Louvain, Belgium.
54 Jorion, P. (1995) 'Predicting Volatility in the Foreign Exchange Market', *Journal of Finance*, Vol. 50, No. 2, pp. 507–528.
55 Latane, H. and Rendleman, R. (1976) 'Standard Deviation of Stock Prices Implied in Option Prices', *Journal of Finance*, Vol. 31, pp. 369–381.
56 Canina, L. and Figlewski, S. (1993) 'The Information Content of Implied Volatility', *Review of Financial Studies*, Vol. 6, pp. 659–681.
57 Neely, C. J. (2003) 'Forecasting Foreign Exchange Volatility: Is Implied Volatility the Best We Can Do?', Federal Reserve Bank of St. Louis Working Paper, 2002–017, St. Louis, MO.
58 As the EUR/USD exchange rate only exists from 4th January, 1999, we follow the approach of Dunis and Williams[95] to apply a synthetic EUR/USD series from 27th November, 1997 to 31st December, 1998 The synthetic series was created using the fixed EUR/DEM conversion rate agreed in 1998, combined with the USD/DEM daily market rate.
59 Baillie, R. T. and Bollerslev, T. (1992) 'Prediction in Dynamic Models with Time Dependent Conditional Variances', *Journal of Econometrics*, Vol. 52, pp. 91–113.
60 Engle, R. F. and Bollerslev, T. (1986) 'Modelling the Persistence of Conditional Variances', *Econometric Review*, Vol. 5, pp. 1–50.
61 Schwert, G. W. (1989) 'Business Cycles, Financial Crises and Stock Volatility', in Brunner, K and Meltzer, A. H. (eds), 'IMF Policy Advice, Market Volatility, Commodity Price Rules and Other Essays', North-Holland, Amsterdam, Netherlands, pp. 82–126.

62 Schwert, G. W. (1989) 'Why Does Stock Market Volatility Change Over Time?', *Journal of Finance*, Vol. 44, pp. 1115–1154.
63 The market data used 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 the analysis owing to the so-called 'smile effect', ie 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.
64 Glossten, C. R., Jagganathan, R. and Runkle, D. E. (1993) 'On the Relation Between the Expected Value and the Volatility of the Nominal Excess Returns on Stocks', *Journal of Finance*, Vol. 48, pp. 1779–1802.
65 Zakoïan, J-M. (1994) 'Threshold Heteroskedasticity Models', *Journal of Economic Dynamics and Control*, Vol. 15, pp. 931–955.
66 Kroner, K. F., Kneafsey, K. P. and Claessens, S. (1995) 'Forecasting Volatility in Commodity Markets, *Journal of Forecasting*, Vol. 14, pp. 77–95.
67 White, H. (1980) 'A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity', *Econometrica*, Vol. 48, pp. 817–838.
68 Billio, M. and Sartore, D. (2003) 'Stochastic Volatility Models: A Survey with Applications to Option Pricing and Value at Risk', in Dunis, C., Laws, J. and Nam, P. (eds), 'Applied Quantitative Methods for Trading and Investment', John Wiley, Chichester, UK, pp. 239–291.
69 Ghysels, E., Harvey, A. C. and Renault, E. (1996) 'Stochastic Volatility', in Maddala, G. S. and Rao, C. R. (eds), 'Handbook of Statistics, 14, Statistical Methods in Finance', North-Holland, Amsterdam, Netherlands, pp. 119–191.
70 Gouriroux, C. and Monfort, A. (1997) 'Time Series and Dynamic Models', Cambridge University Press, Cambridge, UK.
71 Lamoureux, C. G. and Lastrapes, W. D. (1993) 'Forecasting Stock Return Variance: Towards an Understanding of Stochastic Implied Volatilities', *Review of Financial Studies*, Vol. 6, pp. 293–326.
72 Haykin, S. (1999) 'Neural Networks: A Comprehensive Foundation', Prentice-Hall, Englewood Cliffs, NJ.
73 Kaastra, I. and Boyd, M. (1996) 'Designing a Neural Network for Forecasting Financial and Economic Time Series', *Neurocomputing*, Vol. 10, pp. 215–236.
74 Kingdon, J. (1997) 'Intelligent Systems and Financial Forecasting', Springer, London, UK.
75 Deboeck, G. J. (1994) 'Trading on the Edge — Neural, Genetic and Fuzzy Systems for Chaotic Financial Markets', John Wiley, New York, NY.
76 Kuan, C. M. and Liu, T. (1995) 'Forecasting Exchange Rates Using Feedforward and Recurrent Neural Networks', *Journal of Applied Economics*, Vol. 10, pp. 347–364.
77 Franses, P. H. and Van Homelen, P. (1998) 'On Forecasting Exchange Rates Using Neural Networks', *Applied Financial Economics*, Vol. 8, pp. 589–596.
78 Balkin, S. D. (1999) 'Stationarity Concerns When Forecasting Using Neural Networks', Presentation at the INFORMS Conference, 7th–10th November, Philadelphia, PA.
79 Makridakis, S., Andersen, A., Carbone, R. *et al.* (1982) 'The Accuracy of Time Series (Extrapolative) Methods: Results of a Forecasting Competition', *Journal of Forecasting*, Vol. 1, pp. 111–153.
80 Diebold, F. X. and Lopez, J. A. (1996) 'Forecast Evaluation and Combination', *National Bureau of Economic Research*, No. 192, March, NBER, Washington, DC.
81 Makridakis, S., Wheelwright, S. C. and McGee V. E. (1983) 'Forecasting: Methods and Applications', John Wiley, New York, NY.
82 Pindyck, R. S. and Rubinfeld, D. L. (1998) 'Econometric Models and Economic Forecasts', 4th edn, McGraw-Hill, New York, NY.
83 Theil, H. (1966) 'Applied Economic Forecasting', North-Holland, Amsterdam, Netherlands.
84 Basle Committee (1996) 'Amendment to the Capital Accord to Incorporate Market Risks', BIS, Basle, Switzerland.
85 Basle Committee (1996) 'Supervisory Framework for the Use of "Backtesting" in Conjunction with Internal Models Approach to Market Risk', BIS, Basle, Switzerland.
86 Hendricks, D. (1996) 'Evaluation of Value-at-Risk Models Using Historical Data', *Economics Policy Review*, Vol. 2, pp. 36–69.
87 Marshall, C. and Siegel, M. (1997) 'Value-at-Risk Implementing a Risk Measurement Standard', *Journal of Derivatives*, Vol. 4, pp. 91–110.
88 Hull, J. and White, A. (1998) 'Incorporating Volatility Updating into the History Simulation Method for Value-at-Risk', *Journal of Risk*, Vol. 1, pp. 5–19.
89 Jorion, P. (2000) 'Value at Risk: The New

Benchmark for Measuring Financial Risk', 2nd edn, McGraw-Hill, New York, NY.

90 Linsmeier, T. J. and Pearson, N. D. (2000) 'Value at Risk', *Financial Analysts Journal*, Vol. 56, No. 2, pp. 47–66.

91 Janssen, G (2004) 'Advanced Modelling of Conditional Volatility and Correlation in Financial Markets', PhD dissertation, Katholieke Universiteit Leuven, Leuven, Belgium.

92 True, the assumption of normality poses a statistical problem that we are aware of (see Table 2 for the entire dataset), but it is beyond the scope of this paper to benchmark our forecasts against an alternative reference.

93 The volatility filter level is equal to the mean in-sample level of the annualised standard deviation of log-returns plus 20 per cent for both FX series, ie 3.87 per cent for the EUR/USD and 4.76 per cent for the US/JPY.

94 Dunis, C. and Jalilov, J. (2002) 'Neural Network Regression and Alternative Forecasting Techniques for Predicting Financial Variables', *Neural Network World*, Vol. 2, pp. 113–139.

95 Dunis, C. and Williams, M. (2002) 'Modelling and Trading the EUR/USD Exchange Rate: Do Neural Network Models Perform Better?', *Derivatives Use, Trading & Regulation*, Vol. 8, No. 3, pp. 211–239.

96 In fact 20,640 (ie $[223 + (223–4) + (223 - 20)] \times 16 \times 2$).

APPENDIX

Only the summarised out-of-sample results are reported here in order to conserve space. Complete results are available from the authors upon request. Note that scores are rounded up to the closest integer number.

Appendix A1. Overall results for forecasting accuracy

Table A1.1: Overall score of simple models

	EUR/USD vol.		*USD/JPY vol.*	
Simple models	*Average score*	*Overall rank*	*Average score*	*Overall rank*
IM Vol	38	9	37	9
HistoricalVol	10	3	7	1
RM Vol	9	1	10	2
AR_abs	26	6	29	7
AR_sq	28	7	26	6
GARCH_type	23	5	25	5
SV	34	8	32	8
Comb_avg	19	4	21	4
Comb_GR	9	2	11	3

Table A1.2: Overall score of 'mixed' models

'Mixed' models	EUR/USD vol. Average score	Overall rank	USD/JPY vol. Average score	Overall rank
AR_abs	14	4	14	3
AR_sq	22	5	30	7
GARCH_type	27	6	21	5
SV	31	7	27	6
NNR	9	1	11	2
Comb_avg	12	3	15	4
Comb_GR	11	2	6	1

Table A1.3: Average out-of-sample scores for single models vs combination models and simple models vs 'mixed' models

One-day ahead		Five-day ahead		21-day ahead		Overall	
Model	Average score	Model	Average score	Model	Average score	Model	Average score
				EUR/USD			
Simple	22	Simple	22	Simple	22	Simple	22
'Mixed'	18	'Mixed'	18	'Mixed'	18	'Mixed'	18
Single	23	Single	22	Single	22	Single	22
Combination	12	Combination	12	Combination	13	Combination	13
				USD/JPY			
Simple	22	Simple	22	Simple	22	Simple	22
'Mixed'	18	'Mixed'	18	'Mixed'	18	'Mixed'	18
Single	23	Single	22	Single	22	Single	22
Combination	12	Combination	14	Combination	14	Combination	13

Appendix A2. Overall results for VaR efficiency

Table A2.1: Overall score of simple models

Simple models	EUR/USD vol. Average score	Overall rank	USD/JPY vol. Average score	Overall rank
IM Vol	N/A[a]	N/A[a]	N/A[a]	N/A[a]
HistoricalVol	10	=5	11	6
RM Vol	10	=5	12	7
AR_abs	4	2	5	=2
AR_sq	4	3	5	=2
GARCH_type	14	7	8	=4
SV	N/A[a]	N/A[a]	N/A[a]	N/A[a]
Comb_avg	3	1	3	1
Comb_GR	8	4	8	=4

[a]N/A indicates not applicable for ranking.

Table A2.2: Overall score of 'mixed' models

'Mixed' models	EUR/USD vol. Average score	Overall rank	USD/JPY vol. Average score	Overall rank
AR_abs	5	3	5	3
AR_sq	2	1	3	1
GARCH_type	12	6	9	=4
SV	N/A[a]	N/A[a]	N/A[a]	N/A[a]
NNR	6	4	11	6
Comb_avg	9	5	4	2
Comb_GR	5	2	9	=4

[a]N/A indicates not applicable for ranking.

Table A2.3: Average out-of-sample scores for single models vs combination models and simple models vs 'mixed' models

One-day ahead		*Five-day ahead*		*21-day ahead*		*Overall*	
Model	*Score*	*Model*	*Score*	*Model*	*Score*	*Model*	*Score*
				EUR/USD VaR			
Simple	9	Simple	9	Simple	9	Simple	9
'Mixed'	7	'Mixed'	6	'Mixed'	7	'Mixed'	6
Single	7	Single	7	Single	8	Single	7
Combination	6	Combination	6	Combination	6	Combination	6
				USD/JPY VaR			
Simple	9	Simple	9	Simple	9	Simple	9
'Mixed'	7	'Mixed'	7	'Mixed'	7	'Mixed'	7
Single	8	Single	8	Single	8	Single	8
Combination	7	Combination	7	Combination	5	Combination	6

[a]N/A indicates not applicable for ranking.

Appendix A3. Overall results for FX trading performance

Table A3.1: Overall model rankings

EUR/USD Rank	Model	Average score	*USD/JPY* Rank	Model	Average score
	Simple model			**Simple model**	
=6	No vol. filter	32	=6	No vol. filter	29
=3	HistoricalVol	19	=6	HistoricalVol	29
=3	RM Vol	19	=2	RM Vol	20
1	AR_abs	15	5	AR_abs	22
=6	GARCH_type	32	1	GARCH_type	20
5	Comb_avg	29	4	Comb_avg	22
2	Comb_GR	17	=2	Comb_GR	20
	'Mixed' model			**'Mixed' model**	
6	AR_abs	28	5	AR_abs	29
3	AR_sq	20	1	AR_sq	11
=4	GARCH_type	24	3	GARCH_type	20
1	NNR	10	7	SV	35
=4	Comb_avg	24	6	NNR	30
2	Comb_GR	13	2	Comb_avg	14
			4	Comb_GR	25

Table A3.2: Average scores for single models vs combination models and simple models vs 'mixed' models

EUR/USD Model	Average score	*USD/JPY* Model	Average score
Simple	22	Simple	22
'Mixed'	20	'Mixed'	23
Single	21	Single	24
Combination	21	Combination	20