ELSEVIER

Journal of International Financial Markets, Institutions and Money 9 (1999) 407–422

Journal of INTERNATIONAL FINANCIAL MARKETS, INSTITUTIONS & MONEY

www.elsevier.com/locate/econbase

Combining conditional volatility forecasts using neural networks: an application to the EMS exchange rates

Michael Y. Hu [a,*], Christos Tsoukalas [b]

[a] *Graduate School of Management, College of Business Administration, Kent State University, Kent, OH 44242-0001, USA*

[b] *ALLTEL Wholesale Banking Solutions, Inc., 110 East 59th Street, New York, NY 10022, USA*

Received 31 March 1998; accepted 18 April 1999

Abstract

The present paper examines the out-of-sample forecasting performance of four conditional volatility models applied to the European Monetary System (EMS) exchange rates. In order to provide improved volatility forecasts, the four models' forecasts are combined through simple averaging, an ordinary least squares model, and an artificial neural network. The results support the EGARCH specification especially after the foreign exchange crisis of August 1993. The superiority of the EGARCH model is consistent with the nature of the EMS as a managed float regime. The ANN model performed better during the August 1993 crisis especially in terms of root mean absolute prediction error. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Conditional volatility; EMS exchange rates; Neural networks

JEL classification: C32; C45; F31

1. Introduction

The European Monetary System (EMS) as a modern experiment in monetary unification has drawn both criticism and praise since its establishment on March 13, 1979 (Artis, 1990). Founded after the collapse of the Bretton Woods system, the

* Corresponding author. Tel.: +1-330-6722426, ext. 326; fax: +1-330-6722448.
E-mail addresses: mhu@bsa3.kent.edu (M.Y. Hu), ctsoukal@alltelwbs.com (C. Tsoukalas)

PII: S1042-4431(99)00015-3

EMS survived the turbulent 1980s but in the early 1990s faced great challenges when the British pound and the Italian lira dropped out of it on September 17, 1992 and the target zone of $\pm 2.25\%$ ($\pm 6\%$ for the Italian lira) became $\pm 15\%$ on August 2, 1993. Excluding these few periods of high turbulence, the question is whether the EMS has been successful in reducing the volatility of its member currencies.

A number of previous studies have used GARCH models to examine this issue. Diebold and Pauly (1988) used a bivariate ARCH model and detected a structural shift around the time of the establishment of the EMS. Bollerslev (1990) employed a multivariate GARCH model and found reduced volatility and greater coherence for the European exchange rates after March 1979. Nieuwland et al. (1994) provided supporting evidence for the role of the EMS in reducing exchange rate instability. Tsoukalas (1996) employed a multivariate GARCH(1,1) model and showed increased currency stability after the creation of the EMS.

The weakness of most previous studies is their dependence on a single GARCH model that is expected to capture all aspects of the volatility formation process. By combining individual forecasts based on models of different specifications and/or information sets we can better predict future volatility. This is the rationale behind the forecast combining literature which has grown too large to adequately list here. (Literature reviews are provided in Clemen (1989), Granger (1989) and Min and Zellner (1993)) Neural networks, a non-parametric data-driven approach, have been widely used to combine individual forecasts (Donaldson and Kamstra, 1996).

The present paper brings together two strands of research; the conditional volatility models and the neural networks as a forecast combining tool, and sets them in the context of the EMS target zone system. Our objective is to improve on the individual volatility forecasts produced by a number of GARCH models and examine their findings under the light of the EMS. The paper extends previous studies by using a wider range of GARCH models to forecast conditional volatility and combines these models in both parametric (ordinary least squares) and non-parametric (artificial neural network) ways to produce improved volatility forecasts. The forecasting performance of the studied models is evaluated in terms of root mean square and root mean absolute error. The findings point to the superior forecasting performance of the EGARCH model while the results for the artificial neural network are more mixed.

The paper proceeds with Section 2 which presents the research methodology starting with the conditional volatility models and continuing with the forecast combining models one of which is an artificial neural network (ANN). The data set is described in Section 3. The results are presented in Section 4 starting with some diagnostic statistics and proceeding with a number of measures of the models' out-of-sample forecasting performance. The conclusions are drawn in Section 5.

2. Methodology

2.1. Individual volatility forecasts

The GARCH framework (Bollerslev, 1986) is an extension to the basic ARCH model of Engle (1982). Following the tradition of the conditional mean models of the 1970s (e.g. ARMA), the ARCH approach is based on the presumption that forecasts of variance at some future point in time can be improved by using prior information. Despite the lack of a sound economic theory on conditional second moments (with the exception of Hodrick, 1989), these models facilitate the distinction of total volatility into conditional (predictable, and, therefore, less costly to hedge) and unconditional (total) (Artis, 1990). The ARCH specification can explain the observed volatility clustering of exchange rates and provides for normal conditional distributions and symmetric but leptokurtic unconditional ones that closely resemble the empirical distribution of exchange rates (Friedman and Vandersteel, 1982). Previous studies (Bollerslev et al., 1992) have shown that lower-order GARCH models tend to be more parsimonious yet effective in capturing the temporal behavior of volatility. Among the many variants of the original ARCH/GARCH specification in this study we employ the three most popular of them.

The first model is the GARCH(1,1) specification which allows for a longer memory in the conditional variance process than ARCH models. We assume conditional normality and impose non-negativity constraints on the parameters of the variance equation to guarantee positive estimates of the conditional variance. Such constraints sometimes lead to non-convergent solutions.

The main weakness of the simple GARCH model is its assumption of symmetric response to shocks. The exponential GARCH(1,1) (Nelson, 1990a) allows for 'leverage' effects (Black, 1976), that is, for the negative correlation between returns and volatility that is often observed in empirical stock market studies (Christie, 1982). In foreign exchange applications the EGARCH model is useful in capturing asymmetric shocks because it defines conditional variance as an asymmetric function of lagged residuals and it, also, does not require non-negativity constraints.

The third model employed in this study is the integrated GARCH(1,1) model (Engle and Bollerslev, 1986; Nelson, 1990b) which is useful in cases when the volatility has a unit root and a shock to the conditional variance is persistent in the sense that it affects future forecasts of all horizons (integration in variance).

An additional model we used to produce individual volatility forecasts is the MAV model (Pagan and Schwert, 1990) which defines volatility as a simple average of the lagged squared residuals:

$$h_t = \frac{1}{N} \sum_{n-1}^{N} \hat{e}_{t-n}^2 \tag{1}$$

Weighting schemes can be imposed on this process but their infinite sample properties are largely unknown (Bollerslev et al., 1992). The number of lags (N) was set equal to the number of observations over the estimation period.[1]

[1] Previous studies (Donaldson and Kamstra, 1996) used the Schwarz criterion to determine the number of lags 'N'. For the currencies in our data set this criterion gave an average of 1.2 lags which is very close to the lag structure implied by the other conditional volatility models (GARCH, EGARCH, IGARCH). To avoid any overlapping effects we chose a longer lag structure for the MAV model.

2.2. Forecast combining

The individual volatility forecasts produced by these four models were combined through the following formula:

$$F_t = S\left(\beta_0 + \sum_{k=1}^{K} l_k \cdot f_k + \sum_{j=1}^{J} w_{1j} \cdot S\left(\beta_j + \sum_{k=1}^{K} v_{jk} \cdot f_k\right)\right) \tag{2}$$

F_t is the combined volatility forecast for time t, $S(\cdot)$ is a functional form, f_k are the individual forecasts ($K = 4$ in our case), and β_0, β_j, l_k, w_{1j}, and v_{jk} are parameters.

The first forecast combining technique was simple averaging (AVE), that is, estimation of Eq. (2) by setting $w_{1j} = \beta_0 = 0$ and $l_k = 1/4$. The second combining technique was based on ordinary least squares (OLS) to estimate the parameters β_0 and l_k's while all w_{1j}'s are set to 0. The third forecast combining technique used an artificial neural network (ANN).

2.3. Artificial neural networks (ANNs)

Neural networks are non-linear non-parametric models that have their roots in biology and the research on neurons' behavior (Hornik, 1989; Kalman and Kwasny, 1997; Zhang et al., 1998).

The ANN model of this study is a multilayer perceptron (MLP) and consists of one input layer with four input nodes (one for each of the GARCH, EGARCH, IGARCH, and MAV volatility forecasts), one hidden layer with four nodes, and one output layer with one node (the combined forecast). The architecture includes five bias nodes (one for each hidden node and one for the output node) and four skip connections that link the input nodes directly to the output node. This architecture results in a total of 29 parameters: four output-to-hidden-node connections (w_{1j}), 16 hidden-to-input-node connections (v_{jk}), five bias nodes (β_0, β_j), and four skip connections (l_k). The choice of the specific architecture was based on previous empirical evidence that found networks having the number of hidden nodes equal to the number of input nodes to exhibit superior forecasting performance (Tang and Fishwick, 1993). Adding more hidden nodes and/or layers can easily lead to overfitting and poor forecasting performance.

All initial values for the weights and biases were randomly generated from a uniform distribution in the range $[-0.25, 0.25]$ (Weigend et al., 1990). We experimented with several pairs of learning and momentum rates. Small values make the convergence process agonizingly slow while large values lead to oscillation around a local minimum. We ended up with the value of 0.005 which performed the best in our experiments. All inputs to the ANN were linearly normalized to [0,1]. The weights were estimated using the standard backpropagation algorithm and epoch-based training. The actual volatility was approximated by the squared forecast error and was normalized to serve as the target value for the ANN (Donaldson and Kamstra, 1996).

3. Dataset

Our dataset was obtained from Datastream and consists of daily exchange rates covering the 15-year period from the establishment of the European Monetary System on March 13, 1979 to December 30, 1994.[2] The currencies included in the study are those of the first 12 member countries of the European Union (EU), namely, the Belgian/Luxembourg franc (BEF/LUF),[3] the Danish kroner (DKK), the French franc (FRF), the German mark (DEM), the Greek drachma (GRD) the British pound (GBP), the Irish punt (IEP), the Italian lira (ITL), the Dutch guilder (NLG), the Portuguese escudo (PTE), and the Spanish peseta (ESP). The US dollar (USD) was included in the data set as a non-EU currency to serve comparison purposes. All rates are against the German mark because Germany has traditionally been considered the convergence target for all member countries (Fratianni and Von Hagen, 1990).[4] Although not all EU member currencies were officially members of the EMS during the whole study period, we treat them as such because all EU members have the obligation to eventually join the EMS. With a few exceptions (Nieuwland et al., 1994; Tsoukalas, 1996; Hu et al., 1997), most previous EMS studies have ignored the currencies of the smaller EU members (Portugal, Greece). Since both countries are currently in the process of joining the third stage of the monetary unification process, we included them in our data set. Because Belgium and Luxembourg share the same currency and all rates are against the German mark, the data set consists of 11 exchange rates.

The 15-year study period is split up into three subperiods: March 13, 1979–April 4, 1990 (I), April 5, 1990–June 2, 1993 (II), and June 3, 1993–December 30, 1994 (III). Subperiod I is used to estimate the parameters of the conditional volatility models (GARCH, EGARCH, IGARCH, MAV).[5] Then each model produces a one-step-ahead volatility forecast for the first day of subperiod II. The first observation of subperiod I is dropped and the first observation of subperiod II is added in the estimation set and the parameters of the conditional volatility models are estimated again and used to give a one-step-ahead volatility forecast for the 2nd day of subperiod II. This process continues until we get a volatility forecast for each day of subperiods II and III. This 'rolling window' technique has been

[2] Although the date of March 13, 1979 falls within the last period of the dirty float regime, we chose it for this study because it signals the beginning of operation of the EMS. The study period ends in December 1994 for two reasons; firstly, because after that date new member countries entered the EMS (Austria on January 1, 1995 and Finland on October 12, 1996), and, secondly, because we wanted to keep our research setting comparable to that of previous studies (Hu et al., 1997).

[3] The notation used for the currencies is in accordance with ISO 4217.

[4] The use of the US dollar as numeraire currency in previous studies (Bollerslev, 1990) may have induced biases because EMS targets the within-EMS volatility and not the dollar-related volatility.

[5] The parameters of the GARCH-type models were estimated through the Berndt et al. (1974) algorithm.

criticized for 'chasing a moving target' (Refenes et al., 1997). In the EMS case, however, the target (volatility) is supposed to be moving over time (because of the EMS' stabilizing role) and the rolling window technique is indeed appropriate to use.[6]

The forecast combining process worked in a similar way. The parameters of the OLS forecast combining model and the weights of the ANN model were estimated using the volatility forecasts over subperiod II produced as described above. Then, a one-step-ahead combined volatility forecast was produced for the 1st day of subperiod III. Again the rolling window technique is used to successively drop one observation from subperiod II and add one observation from subperiod III to the estimation set. A one-step ahead combined forecast is produced this way for every day of subperiod III.

4. Results

Table 1 Panel A presents some summary statistics on the daily log returns. All mean changes are fairly small and positive indicating depreciation against the German mark over the study period. This finding is additional evidence for the strength of the German mark within the EMS target zone system. The Dutch guilder has the smallest average log return and smallest volatility of all 11 currencies which reflects that country's policy of pegging its currency to the German mark. The US dollar was also depreciated against the German mark during the same period. The Greek drachma has the highest mean log return followed by the Portuguese escudo, the Spanish peseta, and the Italian lira. All exchange rates have positively skewed and leptokurtic distributions which point to significant deviations from normality. Indeed, the Bera and Jarque (1982) joint test of normality confirms that for all rates.

Table 1 also provides information about the autocorrelation structure of the log returns. It presents the first three autocorrelation coefficients (r_1, r_2, r_3) along with their heteroscedasticity-consistent standard errors (Diebold, 1986), the adjusted Box and Pierce statistic (Diebold, 1986) and the Ljung and Box (1978) statistic for the 50th order serial correlation of the original [LB(50)] and the squared log returns [LB2(50)]. McLeod and Li (1983) have shown the Ljung-Box statistic to yield a better fit to the asymptotic chi-squared distribution than the Box-Pierce statistic. The Ljung-Box test rejects the white noise hypothesis for both the original and the squared series of returns. Hsieh (1989) and Bollerslev (1987) interpret the high

[6] Because of the forecasting orientation of this study, a note is needed on the similarity (or lack of it) of the three subperiods. Each of the three subperiods has its share of turbulent events. In terms of exchange rate realignments enforced by the EMS on the member currencies, 15 of them took place in subperiod I, five in subperiod II, and none in subperiod III. Subperiod I includes the stock market crash of 1987, subperiod II the foreign exchange crisis of September 1992, and subperiod III the foreign exchange crisis of August 1993. We expect the rolling window technique to mitigate whatever differences exist among the subperiods.

Table 1
(A) Summary statistics of the daily log returns in the period March 13, 1979–December 30, 1994 (4124 daily observations)[a]. (B) Standard deviations of the daily log returns for each subperiod in the period March 13, 1979–December 30, 1994 (4124 daily observations)[b]

Statistic	Belg./Lux. franc (BEF/LUF)	British, pound (GBP)	Danish kroner (DKK)	Dutch guilder (NLG)	French franc (FRF)	Greek drachma (GRD)	Irish punt (IEP)	Italian lira (ITL)	Port. escudo (PTE)	Spanish peseta (ESP)	US dollar (USD)
	versus German Mark (DEM)										
Panel A											
Mean	0.006	0.011	0.008	0.001	0.010	0.050	0.011	0.020	0.034	0.020	0.004
Standard deviation	0.359	0.475	0.287	0.191	0.237	0.678	0.334	0.358	0.579	0.432	1.066
Coefficient of variation	57.161	43.981	35.207	226.259	24.491	13.532	30.071	17.706	17.254	21.594	245.032
Skewness	0.571	0.629	0.861	0.323	4.681	5.967	3.807	1.773	4.125	1.476	0.679
Kurtosis	14.833	5.533	15.978	48.928	85.800	133.827	71.119	26.748	84.032	43.439	688.971
BJ test	38030**	5531.8**	44380**	411433**	1280032**	3101939**	879079**	125098**	1225067**	325736**	81566382**
r_1	−0.377	0.060	−0.297	−0.390	−0.130	−0.259	−0.255	−0.120	−0.218	−0.160	−0.282
	(0.037)	(0.022)	(0.033)	(0.075)	(0.037)	(0.025)	(0.031)	(0.042)	(0.024)	(0.046)	(0.289)
r_2	−0.010	0.027	−0.037	−0.019	−0.045	−0.028	0.009	−0.012	−0.048	0.035	0.010
	(0.024)	(0.025)	(0.022)	(0.027)	(0.032)	(0.016)	(0.027)	(0.031)	(0.020)	(0.027)	(0.009)
r_3	−0.026	0.011	0.009	−0.034	−0.001	−0.009	0.004	0.011	0.051	−0.004	0.007
	(0.024)	(0.021)	(0.019)	(0.035)	(0.037)	(0.019)	(0.026)	(0.044)	(0.022)	(0.023)	(0.010)
LB(50)	716.711**	117.627**	439.983**	730.700**	159.517**	349.661**	311.463**	148.267**	267.738**	180.372**	356.406**
LB2(50)	731.869**	755.963**	243.464**	987.961**	85.488**	3.603	15.765	953.822**	11.822	233.474**	1025.843**
BP(50)	186.321**	87.326**	145.219**	87.046**	68.641*	185.314**	110.479**	60.095	130.411**	66.785*	42.902

Table 1 (Continued)

Statistic	Belg./Lux. franc (BEF/LUF)	British, pound (GBP)	Danish kroner (DKK)	Dutch guilder (NLG)	French franc (FRF)	Greek drachma (GRD)	Irish punt (IEP)	Italian lira (ITL)	Port. escudo (PTE)	Spanish peseta (ESP)	US dollar (USD)
	versus German Mark (DEM)										
Panel B											
Subperiod I											
March 13, 1979–April 4, 1990	0.376	0.491	0.284	0.221	0.264	0.743	0.311	0.326	0.643	0.428	1.184
Subperiod II											
April 5, 1990–June 2, 1993	0.317	0.461	0.304	0.090	0.122	0.539	0.396	0.435	0.421	0.461	0.774
Subperiod III											
Whole (June 3, 1993–Dec. 30, 1994)	0.315	0.383	0.274	0.073	0.215	0.386	0.355	0.396	0.326	0.403	0.588
Crisis (June 3, 1993–Sept. 30, 1993)	0.585	0.462	0.512	0.079	0.389	0.602	0.445	0.456	0.544	0.708	0.693
Post-crisis (October 1, 1993–Dec. 30, 1994)	0.186	0.360	0.159	0.071	0.136	0.306	0.327	0.379	0.236	0.027	0.558
δ ($\times 10^{-4}$)	−9.95**	−2.635**	−7.116**	−0.133**	−3.790*	−5.396**	−1.768	−2.885*	−7.165**	−11.950**	2.999

[a] BJ test is the Bera and Jarque (1982) joint test of normality that is based on skewness and kurtosis and follows a $x^2(2)$ distribution. r_1, r_2, and r_3 are the coefficients of autocorrelation with 1, 2, and 3 lags (in parentheses are their heteroscedasticity-consistent standard errors according to Diebold, 1986). LB(50) is the Ljung and Box (1978) statistic for the 50th order serial correlation. LB2(50) is the same statistic estimated over the squared log returns of the 11 exchange rates. Both tests follow the x^2 distribution. BP(50) is the adjusted Box and Pierce (Diebold, 1986) statistic for the 50th order serial correlation of the log returns of the 11 exchange rates which also follows the $x^2(50)$ distribution.

[b] δ is the heteroscedasticity-robust regression coefficient of the actual volatility regressed on a daily time variable over subperiod III.

* Statistically significant at the 5% level.

** Statistically significant at the 1% level.

autocorrelation of the squared data as a sign of conditional heteroscedasticity. The strong autocorrelations are consistent with the obligation of the EMS members to keep their currencies within the target zone.

The impact, if any, of the EMS should be reflected on the temporal behavior of the volatility of the exchange rate log returns. Increased monetary cooperation among the member states should lead to less volatile exchange rates. Panel B of the same table shows the volatility pattern of the daily log returns over the whole study period. In all three subperiods the Dutch guilder continues to have the least volatile log returns. The standard deviation of the daily log returns decreased from subperiod I to II for most EMS currencies except for the Danish kroner, the Irish punt, the Italian lira and the Spanish peseta. Certain events account for the volatile behavior of these four currencies. During subperiod II, the Irish punt was devalued once (on February 1, 1993), the Italian lira also once (on September 19, 1992) before eventually exiting the EMS (on September 17, 1992), and the Spanish peseta was devalued twice (on November 23, 1992 and May 14, 1993). The increased volatility of the Danish kroner during subperiod II reflects that country's reaction to the Maastricht Treaty (signed on December 9, 1991) which set up the timetable for the monetary unification process. The Danish reaction was expressed in a referendum that rejected the Maastricht Treaty until a second referendum approved it later on.

Because subperiod III is used to evaluate the out-of-sample performance of the forecasting models, it is appropriate to examine its homogeneity in terms of volatility. Subperiod III includes a period of high turbulence around August 2, 1993 when the EMS target zone of $\pm 2.25\%$ ($\pm 6\%$ for the Italian lira) was widened to $\pm 15\%$. We split up subperiod III into two segments: the 'crisis' period that extends from June 3, 1993 to September 30, 1993, and the 'post-crisis' period that starts on October 1, 1993 and ends on December 30, 1994.[7] Indeed, exchange rate volatility has significantly changed between the crisis and the post-crisis periods. For some currencies the standard deviation was more than three times higher in the crisis period (Belgium/Luxembourg franc, Danish kroner) and for others more than two times higher (French franc, Spanish peseta, Portuguese escudo). The ratio of variance between the crisis and the post-crisis periods is statistically significant at the 1% level for all currencies except for the Dutch guilder that is not significant even at the 5% level ($F_{0.05,85,325} \approx 1.28$ and $F_{0.01,85,325} \approx 1.40$). Also presented in the same table is the heteroscedasticity-robust regression coefficient δ which is estimated by regressing the actual daily volatility on a daily time variable over subperiod III. This coefficient is significantly negative for all EMS currencies except for the Irish punt. Therefore, the volatility seems to have fallen from the peak levels of the crisis period to the lower levels of the post-crisis period.

The forecasting performance of the four individual and the three combining forecasting models was evaluated in terms of two criteria; the root mean square error (RMSE) and the root mean absolute error (RMAE). These are presented in

[7] The cut-off date of October 1 was selected after studying the graphs of the daily log returns for all 11 exchange rates (not shown here).

Table 2
Root mean squared prediction error (RMSE) and root mean absolute prediction error (RMAE) for each exchange rate and volatility forecasting model for subperiod III, June 3, 1993–December 30, 1994 (412 daily observations)[a]

Exchange rate	Forecast error	MAV	GARCH	EGARCH	IGARCH	AVE	OLS	ANN
Belg./Lux. Franc (BEF/LUF)	RMSE	0.533	0.514	0.506	0.524	0.505	0.863	0.552
	RMAE	0.387	0.376	0.354	0.386	0.371	0.439	0.528
British pound (GBP)	RMSE	0.218	0.210	0.213	0.216	0.212	0.224	0.239
	RMAE	0.418	0.388	0.400	0.400	0.400	0.410	0.362
Danish kroner (DKK)	RMSE	0.475	0.465	0.503	0.470	0.466	0.571	0.511
	RMAE	0.327	0.303	0.310	0.315	0.306	0.345	0.511
Dutch guilder (NLG)	RMSE	0.021	0.011	0.008	0.010	0.011	0.009	0.009
	RMAE	0.141	0.089	0.077	0.085	0.098	0.076	0.066
French franc (FRF)	RMSE	0.310	0.325	0.827	0.329	0.391	0.341	0.342
	RMAE	0.238	0.250	0.337	0.253	0.268	0.287	0.427
Greek drachma (GRD)	RMSE	0.445	1.077	0.374	0.406	0.499	0.336	0.380
	RMAE	0.605	1.026	0.437	0.495	0.815	0.417	0.438
Irish punt (IEP)	RMSE	0.274	0.309	0.296	0.299	0.282	0.276	0.283
	RMAE	0.363	0.411	0.391	0.401	0.388	0.378	0.345
Italian lira (ITL)	RMSE	0.303	0.325	0.311	0.315	0.307	0.307	0.306
	RMAE	0.393	0.443	0.427	0.427	0.417	0.437	0.388
Port. escudo (PTE)	RMSE	0.388	0.362	0.342	0.361	0.349	0.383	0.359
	RMAE	0.521	0.418	0.399	0.414	0.435	0.451	0.420
Spanish peseta (ESP)	RMSE	0.506	0.537	0.573	0.550	0.515	0.556	0.518
	RMAE	0.464	0.472	0.467	0.480	0.464	0.504	0.446
US dollar (USD)	RMSE	1.106	2.352	1.056	0.952	1.340	0.627	0.632
	RMAE	1.023	1.513	0.998	0.939	1.133	0.642	0.606

[a] RMSE is the squared root of the mean squared deviation between the conditional volatility predicted by each model and the actual volatility over subperiod III. RMAE is the mean absolute deviation between the conditional volatility predicted by each model and the actual volatility over the same subperiod. MAV is the MA model of volatility which forecasts volatility as the simple average of its own lagged values. GARCH is the simple GARCH(1,1) model of conditional volatility. EGARCH is the exponential GARCH(1,1) model and IGARCH is the integrated-in-variance GARCH(1,1) model. AVE is the model that produces a combined volatility forecast by averaging the forecasts of the four previous models. OLS combines the four models forecasts through an ordinary least squares model. ANN also combines the four models' forecasts through an artificial neural network based on a 4-4-1 architecture, the sigmoid function and the backpropagation algorithm.

Table 2. The ranking of the models in terms of the two criteria is given in Table 3. Panels A and B show the number of times each model (row) got a certain rank (column) in terms of RMSE and RMAE, respectively. Each column adds to 11 which is the number of exchange rates in the data set. The ANN performed better in terms of RMAE (best forecasting model for six currencies) than in terms of RMSE. The fact, however, that the RMSE criterion places more weight on larger forecasting errors undermines the value of ANN as a forecasting tool.

Table 4 presents the best model for each exchange rate in terms of the two criteria over subperiod III. Certain models proved the best under both criteria. Such is the case of the Danish kroner (GARCH for the whole subperiod and EGARCH for the post-crisis period), the French franc (MAV for the whole subperiod), the Greek drachma (OLS for the whole subperiod), the Portuguese escudo (EGARCH for the whole subperiod and for the post-crisis period), and the Spanish peseta (MAV for the whole subperiod). It is evident that during the whole subperiod III no specific model performed the best for all currencies. During the crisis period (June 3, 1993–September 30, 1993) the MAV and ANN models

Table 3
Ranked performance of the forecasting models in the period June 3, 1993–December 30, 1994 (412 daily observations)[a]

Model	Rank						
	1	2	3	4	5	6	7
Panel A: RMSE							
MAV	4	0	0	1	4	0	2
GARCH	2	1	1	1	1	1	4
EGARCH	2	2	1	1	3	0	2
IGARCH	0	0	3	5	1	2	0
AVE	1	4	1	1	1	3	0
OLS	2	2	1	1	0	3	2
ANN	0	2	4	1	1	2	1
Total	11	11	11	11	11	11	11
Panel B: RMAE							
MAV	1	2	1	0	4	0	3
GARCH	1	2	2	0	2	0	4
EGARCH	2	1	3	2	2	1	0
IGARCH	0	1	2	5	1	2	0
AVE	0	3	1	3	1	3	0
OLS	1	2	1	0	1	5	1
ANN	6	0	1	1	0	0	3
Total	11	11	11	11	11	11	11

[a] The ranked performance of each model is estimated as the number of times each model received the specific rank (e.g. 1st, 2nd etc.) in terms of each of the two criteria (RMSE, RMAE). Each column adds up to the number of exchange rates in the study.

Table 4
The best forecasting models in terms of RMSE and RMAE for the whole subperiod III and the crisis and post-crisis periods. Period: June 3, 1993–December 30, 1994 (412 daily observations)[a]

Exchange rate	Forecast error	Best forecasting models		
		Whole period (June 3, 1993–Dec. 30, 1994)	Crisis period (June 3, 1993–Sept. 30, 1993)	Post-crisis period (October 1, 1993–Dec. 30, 1994)
Belg./Lux. franc (BEF/LUF)	RMSE	AVE	AVE	EGARCH
	RMAE	EGARCH	MAV	EGARCH
British pound (GBP)	RMSE	GARCH	MAV	GARCH
	RMAE	ANN	ANN	ANN
Danish kroner (DKK)	RMSE	GARCH	GARCH	EGARCH
	RMAE	GARCH	MAV	EGARCH
Dutch guilder (NLG)	RMSE	EGARCH	EGARCH	EGARCH
	RMAE	ANN	ANN	ANN
French franc (FRF)	RMSE	MAV	ANN	AVE
	RMAE	MAV	MAV	EGARCH
Greek drachma (GRD)	RMSE	OLS	OLS	OLS
	RMAE	OLS	ANN	EGARCH
Irish punt (IEP)	RMSE	MAV	MAV	MAV
	RMAE	ANN	ANN	ANN
Italian lira (ITL)	RMSE	MAV	AVE	MAV
	RMAE	ANN	ANN	ANN
Port. escudo (PTE)	RMSE	EGARCH	MAV	EGARCH
	RMAE	EGARCH	ANN	EGARCH
Spanish peseta (ESP)	RMSE	MAV	MAV	ANN
	RMAE	MAV	ANN	GARCH
US dollar (USD)	RMSE	OLS	OLS	OLS
	RMAE	ANN	ANN	ANN

[a] RMSE is the squared root of the mean squared deviation between the conditional volatility predicted by each model and the actual volatility over subperiod III. RMAE is the mean absolute deviation between the conditional volatility predicted by each model and the actual volatility over the same subperiod. MAV is the MA model of volatility which forecasts volatility as the simple average of its own lagged values. GARCH is the simple GARCH(1,1) model of conditional volatility. EGARCH is the exponential GARCH(1,1) model and IGARCH is the integrated-in-variance GARCH(1,1) model. AVE is the model that produces a combined volatility forecast by averaging the forecasts of the four previous models. OLS combines the four models forecasts through an ordinary least squares model. ANN also combines the four models' forecasts through an artificial neural network based on a 4-4-1 architecture, the sigmoid function and the backpropagation algorithm.

performed adequately well. In the post-crisis period (October 1, 1993–December 30, 1994), the EGARCH model performed the best for many currencies. What is needed is a metric to capture the overall performance of a model across currencies. Such a metric is the 'rank sum' (Hippel and McLeod, 1993).

Table 5 presents the rank sum for each model, which is defined as the sum of the product of each rank and the number of times the model received that rank. Models with low rank sums forecast better overall than models with higher rank sums. During the crisis (June 3, 1993–September 30, 1993) the best model in terms

Table 5
Ranked performance of the forecasting models in terms of the rank sums for the whole subperiod III, the crisis and the post-crisis periods. Period: June 3, 1993-December 30, 1994 (412 daily observations)[a]

Model	Time periods					
	Whole period (June 3, 1993–Dec.30, 1994)		Crisis period (June 3, 1993–Sept. 30, 1993)		Post-crisis period (October 1, 1993–Dec. 30, 1994)	
	Rank sum	Ranking	Rank sum	Ranking	Rank sum	Ranking
Panel A: RMSE						
MAV	42	2	35	1	54	7
GARCH	50	7	54	7	46	5
EGARCH	42	3	44	4	32	1
IGARCH	46	6	52	6	45	4
AVE	39	1	38	2	40	2
OLS	45	5	46	5	44	3
ANN	44	4	39	3	47	6
Panel B: RMAE						
MAV	49	5	29	2	58	7
GARCH	49	6	60	7	45	4
EGARCH	37	2	45	3	26	1
IGARCH	45	4	58	6	37	2
AVE	44	3	48	5	50	5
OLS	50	7	47	4	52	6
ANN	34	1	21	1	40	3

[a] Rank sum is the sum of ranks (e.g. 1st, 2nd etc.) that each model received for each of the 11 currencies. The smaller the rank sum is, the better the forecasting performance of the model across the 11 currencies and the higher its ranking. RMSE is the squared root of the mean squared deviation between the conditional volatility predicted by each model and the actual volatility over subperiod III. RMAE is the mean absolute deviation between the conditional volatility predicted by each model and the actual volatility over the same subperiod. MAV is the MA model of volatility which forecasts volatility as the simple average of its own lagged values. GARCH is the simple GARCH(1,1) model of conditional volatility. EGARCH is the exponential GARCH(1,1) model and IGARCH is the integrated-in-variance GARCH(1,1) model. AVE is the model that produces a combined volatility forecast by averaging the forecasts of the four previous models. OLS combines the four models' forecasts through an ordinary least squares model. ANN also combines the four models' forecasts through an artificial neural network based on a 4-4-1 architecture, the sigmoid function and the backpropagation algorithm.

of RMSE is the MAV followed by the AVE and in terms of RMAE is the ANN followed by the MAV. After the crisis (October 1, 1993–December 30, 1994) the best model in terms of RMSE is the EGARCH followed by the AVE and in terms of RMAE is the EGARCH followed by the IGARCH.[8]

The findings from these tables can be summarized as follows. First, the EGARCH model has a superior out-of-sample forecasting performance in the post-crisis period. Following the widening of the target zone to $\pm 15\%$ there was a period of high volatility which for most currencies ended around the end of September 1993. During that period, many member countries continued to intervene in the foreign exchange markets to keep their exchange rates under control. Such interventions, assisted by credit lines provided through the EMS, created higher volatility and asymmetric effects since the interventions depended not only on the magnitude but also on the direction of the shock. One way to capture such asymmetric effects is to examine the correlation coefficient between the daily log returns and the actual volatility. These correlation coefficients were positive for all currencies in the crisis period (ranging from 0.073 for the Dutch guilder to 1.155 for the Belgian/Luxembourg franc) but became negative in the post-crisis period (ranging from -0.001 for the Irish punt to -0.457 for the Belgium/Luxembourg franc). In fact, they are negative for all currencies whose best performing model in the post-crisis period is the EGARCH (Table 4) and positive for all currencies whose best performing model is not the EGARCH under any of the two criteria (i.e. Italian lira, Spanish peseta, US dollar).

Second, the MAV model has a satisfactory out-of-sample forecasting performance especially in the crisis period. This may be due to the longer memory incorporated in the MAV model compared to the GARCH type of models. Given the statistically significant autocorrelation structure in the raw data reported here and in previous studies (Hu et al., 1997), more attention should be paid to AR(k)-GARCH(p,q) specifications.

Third, the ANN forecast combining model performed better in the crisis period and it proved superior to linear combining models (e.g. AVE, OLS). Its performance could improve further by selecting a more appropriate architecture for each currency and period (crisis, post-crisis). The optimum ANN architecture is dependent on the volatility characteristics of each exchange rate. Expressing the ANN model as a function of the exchange rate characteristics can facilitate the forecasting effort by providing guidelines to be used in selecting networks.

Finally, there are differences among the different performance measures. Certain models performed better in terms of RMSE than in terms of RMAE. In a real business setting the appropriate criterion will be selected after accounting for the cost of forecasting errors.

[8] The IGARCH model was not the best model for any currency (Table 4) but it was second or third best for many of them (not shown here).

5. Conclusions

The present study examined the out-of-sample forecasting performance of a number of conditional volatility models for a set of 11 currencies and a period covering the first 15 years of operation of the EMS. The empirical findings provide evidence for the positive role of the EMS in reducing the volatility of the member currencies despite the foreign exchange crises that have plagued the system. Empirical evidence is also provided for the superior out-of-sample forecasting performance of the EGARCH model which is due to the monetary arrangements enforced by the EMS.

The ability of the MAV and ANN models to account for the observed volatility during the foreign exchange crisis of August 1993 is an interesting issue deserving further investigation. Probably, models with longer memory are more appropriate to account for the autocorrelation structure generated in crisis periods. The assumption of conditional normality does not seem to describe the data well. The remedy for the persistent deviations from normality may be found in the use of highly skewed distributions or mixtures of distributions.

This study also provides evidence for the forecast combining performance of the ANN model. It seems to behave well in the crisis period and gets a good ranking (no. 3) in the post-crisis period. Its performance is much better in terms of absolute prediction errors than in terms of squared prediction errors.

An issue raised by this study is the relationship between exchange rate pegging and volatility across subperiods (most notable is the case of the Dutch guilder). This issue was not addressed in the present study but deserves further investigation. Future research should also examine the temporal pattern of the connection weights of the ANN model to determine how each input (volatility forecast) contributes to the final combined forecast over time. Such issues open up whole new avenues of research on the application of neural networks in finance.

References

Artis, M.J., 1990. The European Monetary System: a review of the research record. In: Bird, G. (Ed.), The International Financial Regime. Surrey University Press, UK.

Bera, A.K., Jarque, C.M., 1982. Model specification tests: a simultaneous approach. J. Econometrics 20, 59–82.

Berndt, E.K., Bronwyn, H.H., Hall, R.E., Hausman, J.A., 1974. Estimation and inference in non-linear structural models. Ann. Econ. Soc. Meas. 4, 653–665.

Black, F., 1976. Studies of stock price volatility changes. In: Proceedings from the American Statistical Association Annual Conference, Business and Economic Statistics Section. American Statistical Association, Washington DC, pp. 177–181.

Bollerslev, T., 1990. Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. Rev. Econ. Statistics 72, 498–505.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroscedasticity. J. Econometrics 31, 307–327.

Bollerslev, T., 1987. A conditionally heteroscedastic time series model for speculative prices and rates of return. Rev. Econ. Statistics 9, 542–547.

Bollerslev, T., Chou, R.Y., Kroner, K.F., 1992. ARCH modeling in finance. J. Econometrics 52, 5–59.
Christie, A.A., 1982. The stochastic behavior of common stock variances: value, leverage and interest rate effects. J. Financial Econ. 10, 407–432.
Clemen, R.T., 1989. Combing forecasts: a review and annotated bibliography. Int. J. Forecasting 5, 559–583.
Diebold, F.X., 1986. Testing for serial correlation in the presence of ARCH. Unpublished manuscript, University of Pennsylvania.
Diebold, F.X., Pauly, P., 1988. Has the EMS reduced member-country exchange rate volatility? Empirical Econ. 13, 81–102.
Donaldson, G.R., Kamstra, M., 1996. Forecast combining with neural networks. J. Forecasting 15, 49–61.
Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. Econometrica 50, 987–1008.
Engle, R.F., Bollerslev, T., 1986. Modelling the persistence of conditional variances. Econometric Rev. 5, 1–87.
Fratianni, M., Von Hagen, J., 1990. German dominance in the EMS: the empirical evidence. Open Economies Rev. 1, 86–87.
Friedman, D., Vandersteel, S., 1982. Short-run fluctuations in foreign exchange rates. J. Int. Econ. 13, 171–186.
Granger, C.W.J., 1989. Invited review: combining forecasts—20 years later. J. Forecasting 8, 167–173.
Hippel, K.W., McLeod, A.L., 1993. Time-Series Modelling of Water Resources and Environmental Systems. Elsevier, Amsterdam.
Hodrick, R.J., 1989. Risk, uncertainty, and exchange rates. J. Monetary Econ. 23, 433–459.
Hornik, K., 1989. Multilayer feedforward networks are universal approximators. Neural Networks 2, 359–366.
Hsieh, D.A., 1989. Modeling heteroscedasticity in daily foreign-exchange rates. J. Bus. Econ. Statistics 7, 307–317.
Hu, M.Y., Jiang, C.X., Tsoukalas, C., 1997. The conditional behavior of the European exchange rates before and after the establishment of the European Monetary System. J. Int. Financial Markets Institutions Money 7, 235–253.
Kalman, B.L., Kwasny, S.C., 1997. High performance training of feedforward and simple recurrent networks. Neurocomputing 14, 63–83.
Ljung, G., Box, G., 1978. On a measure of lack of fit in time-series models. Biometrika 65, 297–303.
McLeod, A.I., Li, W.K., 1983. Diagnostic checking ARMA time series models using squared-residual autocorrelations. J. Time Series Anal. 4, 269–273.
Min, C., Zellner, A., 1993. Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates. J. Econometrics 56, 89–118.
Nelson, D.B., 1990a. Conditional heteroscedasticity in asset returns: a new approach. Econometrica 59, 347–370.
Nelson, D.B., 1990b. Stationarity and persistence in the GARCH(1,1) model. Econometric Theory 6, 318–334.
Nieuwland, F.G.M.C., Veshoor, W.F.C., Wolff, C.C.P., 1994. Stochastic trends and jumps in EMS exchange rates. J. Int. Money Finance 13, 699–727.
Pagan, A.R, Schwert, G.W., 1990. Alternative models for conditional stock volatility. J. Econometrics 45, 267–290.
Refenes, A.P., Bentz, Y., Burgess, N.A., Bunn, A.N., Zapranis, A.D., 1997. Financial time series modelling with discounted least squares backpropagation. Neurocomputing 14, 123–138.
Tang, Z.P., Fishwick, A., 1993. Feedforward neural nets as models for time series forecasting. ORSA J. Computing 5, 374–385.
Tsoukalas, C., 1996. The volatility impact of the European Monetary System on member and non-member currencies. Unpublished doctoral dissertation, Graduate School of Management, Kent State University.
Weigend, A.S., Rumelhart, D.E., Huberman, B.A., 1990. Predicting the future: a connectionist approach. Int. J. Neural Systems 1, 193–209.
Zhang, G., Patuwo, E.B., Hu, M.Y., 1998. Forecasting with artificial neural networks: the state of the art. Int. J. Forecasting 14, 35–62.