CRENoS

Centro Ricerche Economiche Nord Sud
Università degli Studi di Cagliari

EVALUATING NON-LINEAR MODELS ON *POINT* AND *INTERVAL* FORECASTS: AN APPLICATION WITH EXCHANGE RATE RETURNS

Gianna Boero
Emanuela Marrocu

CONTRIBUTI DI RICERCA
01/10

Gianna Boero
University of Cagliari, CRENoS and University of Warwick
e-mail: Boero@unica.it

Emanuela Marrocu
University of Cagliari and CRENoS
e-mail: emarrocu@unica.it

EVALUATING NON-LINEAR MODELS ON *POINT* AND *INTERVAL* FORECASTS:AN APPLICATION WITH EXCHANGE RATE RETURNS

Abstract

The aim of this paper is to compare the forecasting performance of SETAR and GARCH models against a linear benchmark using historical data for the returns of the Japanese yen/US dollar exchange rate. The relative performance of the models is evaluated on point forecasts and on interval forecasts. Point forecasts evaluation over the whole forecast period indicates that the performance of the models, when distinguishable, tends to favour the linear models. However, we show that if the evaluation of point forecasts is conducted over distinct sub-samples or specific regimes there is more evidence of forecasting gains, especially from the SETAR models. Moreover, when we evaluate the validity of interval forecasts, the results produce clear evidence of the superiority of the non-linear models, and tend to favour especially the GARCH models.

JEL: C22, C51, C53, E17
Keywords: nonlinearity, asymmetry, forecasting accuracy, point forecasts, interval forecasts, exchange rates.

December 2001

1. Introduction

Several studies have been conducted in the context of univariate models, exploiting recent developments of nonlinear time series econometrics. These studies, theoretically based on the efficient market hypothesis, are mainly focused on the dynamic representation of exchange rates and on their short run predictability. The main rationale for these models is that if the exchange rates market is characterised by a certain degree of efficiency, it is plausible to assume that all the relevant information is embodied in the most recent exchange rates returns, so that it becomes unnecessary to include the economic fundamentals in the set of explanatory variables. Among the most commonly applied nonlinear models, the GARCH (generalised autoregressive conditional heteroscedastic) and the SETAR (self-exciting threshold autoregressive) models have proved successful in describing the dynamic behaviour of many economic and financial variables. The GARCH models allow one to specify the process governing both the mean and the variance of the series, while the SETAR models represent a stochastic process generated by the alternation of different regimes. Moreover, the GARCH model is particularly suitable to describe the typical behaviour of financial time series, namely the fact that large (small) price changes tend to be followed by large (small) price changes of either sign; this kind of dependency can be mostly exploited to improve interval forecasts. An improvement in point forecasts can be achieved by the GARCH in Mean (GARCH-M) model, where the conditional variance estimate enters as a regressor in the mean equation of the series. Many authors have also stressed the empirical relevance of nonlinearity in mean for the exchange rate returns; we refer, among others, to Meese and Roose (1991), Kräger and Kugler (1993), Peel and Speight (1994) and Brooks (1997).

Although there have been extensive applications of new techniques to describe the nonlinearities and asymmetries which characterise exchange rate dynamics, there are still few studies on the forecasting performance of the different models for historical time series data; comparisons have been carried out typically with respect to the random walk model or, more recently, by means of simulated data based on Monte Carlo experiments (see, for example, Clements and Smith, 1997, 1999). In general, the significant presence of mean-nonlinearities for the in-sample period only rarely has provided better out-of-sample forecasts compared with those obtained from a simple linear or a random walk model. Furthermore, the results are often sensitive to the length of the forecast horizon and to the metric adopted to measure the forecasting accuracy.

Diebold and Nason (1990) suggest four different reasons why nonlinear models cannot provide better out-of-sample forecasts than the simpler linear model even when linearity is significantly rejected in-sample: 1) nonlinearities concern the even-ordered conditional moments and therefore are not useful for improving forecasts; 2) in-sample nonlinearities are due to structural breaks or outliers which cannot be exploited to improve out-of-sample forecasts; 3) conditional means nonlinearities are a feature of the DGP but are not large enough to offer better forecasts; 4) nonlinearities are present but they are captured by the wrong type of nonlinear model.

Dacco and Satchell (1999) and Clements and Smith (2001) argue that the alleged *poor* forecasting performance of nonlinear models can also be due to the evaluation and measurement method adopted. Clements and Smith show, on the basis of a Monte Carlo study, that the evaluation of the whole forecast density may reveal gains to the nonlinear models which are systematically masked if the comparison is carried out only in terms of MSFE. This result has been confirmed by Boero and Marrocu (2001) in an application with actual data. Dacco and Satchell (1999) suggest that methods based on the profitability criterion should turn out to be more adequate in the case of financial variables; thus, tests for the percentage of correct sign predictions, are expected to be more informative in deciding whether to buy or sell foreign currencies. Contrary to this suggestion, the study by Boero and Marrocu

(2001) did not show significant evidence in favour of the nonlinear models on the basis of the percentage of correct sign predictions. The aim of this paper is to compare the forecasting performance of SETAR and GARCH models against the AR benchmark using weekly data for the returns of the Japanese yen against the US dollar[1]. We conduct the forecast evaluation using different criteria, namely *point forecasts* and *interval forecasts*. The measure adopted for the evaluation of point forecasts is the mean square forecast error (MSFE). Interval forecasts are evaluated by means of the LR tests of correct conditional coverage as recently proposed by Christoffersen (1998). We start the analysis by comparing MSFEs from the competing models over the entire forecasting period. We then evaluate the models over different out-of-sample periods obtained by splitting the initial forecasting sample in six sub-samples of equal length; this exercise should reveal potential gains from non-linear models in periods characterised by strong nonlinear features. We also investigate the possibility that nonlinear models are more valuable in terms of forecasting accuracy when the process is in a particular regime, for example depreciation or appreciation. Next, we supplement the analysis of point forecasts by constructing and evaluating interval forecasts. The use of interval forecasts is becoming increasingly common in practical real-life applications, as they provide a description of forecast uncertainty which is not available from point forecasts alone. Models of conditional variance such as GARCH are particularly suitable to provide some indication of the uncertainty around the forecast, and could therefore exhibit accuracy gains, when evaluated on interval forecasts, which may be systematically masked in MSFE comparisons.

The rest of the paper is organised as follows. In section 2 we describe the models adopted. In section 3 we present the statistical properties of the data and the results of the tests performed to

[1] For a study of the forecast performance of exchange rate returns at different frequencies (monthly and daily) see Boero and Marrocu, 2000.

detect the presence of non-linearities. The findings from the modelling and forecasting exercises are reported in sections 4 and 5, respectively. Finally, in section 6 we summarise the main results and make some concluding remarks.

2. The models

2.1 The threshold autoregressive models

Threshold autoregressive models were first proposed by Tong (1978), Tong and Lim (1980) and Tong (1983). The essential idea of this class of nonlinear model is that the behaviour of a process can be described by a finite set of linear autoregressions. The appropriate AR model that generates the value of the time series at each point in time is determined by the relation of a conditioning variable to the threshold values; if the conditioning variable is the dependent variable itself after some delay, *d*, the model is known as *self-exciting*, hence the acronym SETAR.

The SETAR model is piecewise-linear in the space of the threshold variable, rather than in time. If the process is in the j^{th} regime, the p^{th} order linear autoregression is formally defined as:

$$y_t = \phi_0^{(j)} + \phi_1^{(j)} y_{t-1} + \ldots + \phi_p^{(j)} y_{t-p} + \varepsilon_t^{(j)} \qquad \text{for } r_{j-1} \le y_{t-d} < r_j \qquad (1)$$

$$\varepsilon_t^{(j)} \sim IID(0, \sigma^{2(j)})$$

In order to allow for different autoregressive structures across regimes, *p* can be seen as the maximum lag order. An interesting feature of SETAR models is that the stationarity of y_t does not require the model to be stationary in each regime, on the contrary the limit cycle behaviour that this class of models is able to describe arises from the alternation of explosive and contractionary regimes.

A variant of the TAR model can be obtained if the parameters are allowed to change smoothly over time, the resulting model is called a Smooth Transition Autoregressive (STAR) model (see Granger and Teräsvirta , 1993, and Teräsvirta, 1994).

When the structural parameters, *r* and *d*, are known a SETAR model can be estimated by fitting an AR model to the appropriate subset of observations determined by the relationship of the threshold variable to the value of the threshold (*arranged autoregression*).

In the case in which the threshold parameter (r) and the delay parameter (d) are unknown, Tong (1983) suggests an empirical procedure that allows selecting as the best model the one which yields the minimum Akaike Information Criteria (AIC). However, as stressed by Priestley (1988), such a procedure has to be seen as a guide in choosing a small subclass of nonlinear models featuring desirable economic and statistical properties.

For the case of a SETAR (p_1, p_2; d) model Tong (1983) proposes a three-stage procedure: for given values of d and r, separate AR models are fitted to the appropriate subsets of data, the order of each model is chosen according to the usual AIC criteria. In the second stage r can vary over a set of possible values while d has to remain fixed, the re-estimation of the separate AR models allows the determination of the r parameter, as the one for which AIC(d) attains its minimum value. In stage three the search over d is carried out by repeating both stage 1 and stage 2 for $d=d_1$, d_2, ..., d_p. The selected value of d is, again, the value which minimise AIC(d).

2.2 GARCH models

An ARCH process can be defined in terms of the error distribution of a model in which the variable y_t is generated by:

$$y_t = x_t b + e_t \qquad t=1,...,T \tag{2}$$

where x_t is a vector of $kx1$ explanatory variables, which in our study includes only lagged values of y_t, and â is a $kx1$ vector of autoregressive coefficients. The ARCH model proposed by Engle (1982) specifies the distribution of $å_t$ conditioned on the information set Y_{t-1}, which includes the actual values for the variables y_{t-1}, y_{t-2}, . y_{t-k}. In particular, the model is based on the assumption that:

$$e_t \mid \Psi_{t-1} \sim N(0,h_t)$$

where $h_t = a_0 + a_1 e_{t-1}^2 + ... + a_q e_{t-q}^2$ (3)

with á$_0$>*0* and á$_i$³*0*, *i=1, ..., q*, in order to constrain the conditional variance to be positive. Thus, the error variance is time-varying and depends on the magnitude of past errors.

Bollerslev (1986) proposes a generalisation of the ARCH model, which leads to the following specification of the conditional variance:

$$h_t = a_0 + a_1 e_{t-1}^2 + ... + a_q e_{t-q}^2 + b_1 h_{t-1} + ... + b_p h_{t-p} \quad (4)$$

This process is known as GARCH(*p*,*q*). To guarantee that the conditional variance assumes only positive values the following restrictions have to be imposed: á$_0$>*0*, á$_i$³*0* for *i=1,...,q*, and â$_i$³*0* for *i=1, ..., p*. In practice, the value of *q* in the GARCH model is much smaller than the corresponding value of *q* in the ARCH representation. Usually, a simple GARCH(1,1) model offers an adequate description of most economic and financial time series.

Engle, Lilien and Robins (1987) extend the ARCH model by introducing the conditional variance as a regressor in the mean equation of the variable:

$$y_t = x_t^{'} b + d h_t + e_t \qquad t=1,...,T \quad (5)$$

where $e_t | \Psi_{t-1}$ ~N(*0*,h_t) and h_t is a (G)ARCH process. In the (G)ARCH-M model the conditional variance is included in the mean equation according to different functional forms: $\log(h_t)$, $\sqrt{h_t}$ and h_t.

A relevant extension of the GARCH models is represented by the class of asymmetric models. These models allow one to capture possible asymmetries in the conditional variance induced by the sign and the magnitude of past shocks. Most applied specifications are the threshold heteroscedastic model, (TARCH) (Glosten, Jagannathan and Runkle, 1993, and Zakoian, 1994), and the Exponential GARCH model (EGARCH, Nelson, 1991).

3. Preliminary data analysis and linearity tests

The empirical analysis has been carried out on the exchange rate returns measured in log-differences. The log-levels and the returns of the series for the period 1973.1-1997.7 (1281 observations) are depicted in Figure 1. The returns series is mean-stationary, while the variance features the typical *volatility clustering* phenomenon with periods of high volatility followed by periods of low volatility. Table 1 reports the summary of the descriptive statistics for the exchange rate returns. The series is characterised by excessive kurtosis and asymmetry; the Jarque-Bera test strongly rejects the normality hypothesis.

In order to detect the presence of nonlinear components in the returns series we apply the RESET test and the S_2 test proposed by Luukkonen-Saikkonen-Teräsvirta (1988). These tests are devised for the null hypothesis of linearity. The RESET test is applied in its Lagrange Multiplier variant (Granger-Teräsvirta, 1993): a linear autoregression of order p is run, followed by an auxiliary regression in which powers of the fitted values obtained in the first stage are included along with the initial regressors up to the power h = 2, 3, 4. The test is distributed as a $\div^2$ with h-1 degrees of freedom. While the RESET test is devised for a generic form of misspecification, the S_2 test is formulated for a specific alternative hypothesis, i.e. STAR-type nonlinearity; the authors show that the S_2 test has reasonable power even when the true model is a SETAR one. The test is calculated as $S_2=T(SSE_0-SSE_1)/SSE_0$, where SSE_0 is the residual sum of squares from a linear autoregression of order p for y_t, and SSE_1 is the residual sum of squares from the auxiliary regression in which the initial regressors enter linearly and multiplied by the transition variable y_{t-d} raised up to the third power[2]. In this analysis we perform the test

[2] The auxiliary regression is specified as:

$$\hat{e}_t = b_0 + \sum_{i=1}^{p} b_i y_{t-i} + \sum_{i=1}^{p} x_i y_{t-i} y_{t-d} + \sum_{i=1}^{p} y_i y_{t-i} y_{t-d}^2 + \sum_{i=1}^{p} k_i y_{t-i} y_{t-d}^3 + n_t,$$

where $\hat{e}_t$ are the residuals from the linear AR(p) model.

selecting the value of the delay parameter, *d*, in the range [1,5]; under the null hypothesis of linearity the test has a χ^2 distribution with 3*p* degrees of freedom.

In Table 2 we report the probability values for the tests computed for the whole sample period (n=1281), the estimation period (n=964) and the forecast period (n=313). For each test the linear model under the null hypothesis has been estimated assuming different lag structures (*p*=2, , 6). The table reports results only for *p*=3, 4, and 5. As we can see, when the tests are applied to the whole sample, they lead to the rejection of the null in a large number of cases, indicating that there is strong evidence of nonlinear components in the data. However, by splitting the sample into the estimation period and the forecast period we notice that there is less evidence of non-linearities in the latter. When the tests are applied to the forecast period, in fact, we obtain clear evidence of nonlinearity only from the S_2 test with *d*=1. These results will be taken into account later on, in the evaluation of the forecast performance of the models.

4. models estimation

The models were estimated over the period 1973.2-1991.6. Examination of the in-sample returns data revealed some significant serial correlation at lag 2. We therefore selected a restricted AR(2) model for the Japanese yen exchange rate returns. With regard to the SETAR models, we estimated specifications with one-threshold (2 regimes) and two-thresholds (3 regimes), following the estimation procedure suggested by Tong (1983).[3] The model selection has been conducted on the basis of the AIC criterion; however, when it appeared that the AIC overestimated the autoregressive order of the model, we selected the model with the most parsimonious dynamic structure. Moreover, we considered only models with a maximum lag order *p*=6. The

[3]All the models have been selected and estimated with Eviews codes; the codes are available from the authors upon request.

models selected are reported in Table 3[4]. In general, the dynamic structure, the estimated coefficients and the error variance differ significantly across regimes, therefore indicating that the data are strongly characterised by non-linearities. Moreover, it is interesting to note that the dynamics of the three-regime SETAR model are in line with the theoretical model described in Hsieh (1989) and with the empirical evidence reported by Kräger and Kugler (1993): the evidence of non-linearity in the mean is likely to be due to the existence of a managed floating exchange rate regime, in which the central banks intervene in order to avoid excessive depreciation or appreciation.

GARCH components were strongly present in the data, thus capturing the evident volatility clustering illustrated in Figure 1. In order to describe appropriately such components, we identified some alternative models, namely, a simple GARCH(1,1), an EGARCH(1,1) and a TARCH(1,1) to take into account possible asymmetries in the conditional variance, and a GARCH in mean (GARCH-M(1,1)). The *best* model was selected according to the Akaike (AIC) and Schwarz (SIC) information criteria. A marginally significant variance component was found in the mean equation for the returns of the Japanese yen; the resulting joint estimated AR model for the mean of the returns was of a lower order (AR(1) process) than that reported under AR estimation alone.

5. The forecasting exercise

The forecasting performance of the models is evaluated in different ways. First we compute MSFE for the various models for different steps ahead (1 to 5), and compare the relative performance of the models by means of the Diebold and Mariano test. This exercise is first conducted over the entire period, then on different sub-samples where nonlinearities may be present with

[4] Note that even the simplest SETAR models with an AR(1) process in each regime can generate complex dynamic behaviour. Moreover, it is worthwhile stressing that the constant terms plays a relevant role in nonlinear models.

different intensities. Second, following other authors (Tiao and Tsay, 1994 and Clements and Smith, 2001), we analyse the forecast performance of the models conditional on the regime of the SETAR models, to see whether these models show a better performance for observations following in a specific regime. Finally, we extend the evaluation of the models to cover interval predictions. Models of conditional variance such as GARCH are mostly useful when the object of the analysis is to provide some indication of the uncertainty around the mean. Evaluation of interval forecasts could reveal gains to the non-linear models, particularly the GARCH models, which may not be apparent on MSFE measures.

5.1. Point forecasts evaluation

5.1.1 MSFEs over the entire forecast period

In this comparative exercise the forecasting ability of the models is assessed by means of the MSFE. The forecasts of the yen exchange rate returns have been calculated recursively from 1 to 5 steps-ahead. The models were identified and specified only once, over the first estimation periods, 1973.2-1991.6. The models were then re-estimated (but not re-specified) by expanding the sample with one observation each time, over the period 1991.7-1997.7, therefore obtaining 313 point forecasts for each forecasting horizon (*h*). The computation of multi-step-ahead forecasts (*h*>1) from non-linear models (SETAR) involves the solution of complex analytical calculations and the use of numerical integration techniques, or alternatively, the use of simulation methods. In this study the forecasts are obtained by applying the Monte Carlo[5] method, following the suggestions in Clements and Smith (1997, 1999). In Table 4 we report the MSFEs and MSFEs normalised with respect to the linear model, which represents our benchmark. The values are calculated as the ratio $MSFE_{NL}/MSFE_{L}$; a number less than one means that the non-

[5] Each point forecast is obtained as the average over 500 replications.

linear model provides more accurate forecasts than the simple linear model. Furthermore, in order to assess whether such a superiority is statistically significant we perform the Diebold-Mariano (DM, 1995) test; values leading to the rejection of the null hypothesis of equality of forecast accuracy are indicated with stars. Table 4 also reports the MSFE obtained from a *naïve* forecast by assuming that the levels of the exchange rates follow a *random walk* with drift process.

We note that in terms of MSFE the models exhibit in general similar values. The performance of the models is significantly different in three cases out of the fifteen considered, with only one case in favour of the non-linear model: the SETAR-2 dominates the linear model in the 5 steps ahead. Diebold and Nason (1990) argue that non-linearity may not be pronounced enough over the whole forecast period to guarantee greater forecast accuracy of the non-linear models. As we have seen from the results in Table 2, the linearity tests showed weaker evidence of non-linearities over the forecast period than in the estimation period, with the exception of the S_2 test with d=1. This may explain the results in Table 4.

Other explanations for the inadequate forecasting performance of the nonlinear models have been offered: Clements and Smith (2001), for example, have shown that forecasting gains of nonlinear models may be masked by the evaluation method adopted or may depend on the *state of nature*.

In what follows, by pursuing the explanation in Diebold and Nason, we examine whether the results reported in Table 4 are due to the weaker presence of non-linearity features in the forecast period. We divide the forecast sample in six sub-periods of equal length, and conduct the linearity test in each sub-sample, in order to detect possible varying degrees of non-linearity. We then compare the models on the basis of the MSFEs obtained over the different sub-samples, expecting the SETAR and GARCH models to offer greater forecasting gains in sub-periods with stronger non-linearities. In a second exercise, we condition the forecast observations on the regimes of the SETAR models and examine whether the SETAR forecasts are superior to the linear forecasts conditional on a specific regime. Tiao and Tsay (1994) for example, have shown that SETAR models produce US GNP forecasts which are superior to those obtained from a linear model, when the forecasts are obtained from the regime with fewer observations (when the economy is recovering from recession). Similarly, Clements and Smith (2001) have shown, by means of Monte Carlo simulations, that a 3-regimes SETAR model for the yen exchange rate returns records significant gains (over 40%) relative to the random walk model for the one-step-ahead forecasts conditional on being in the Middle regime. Like in the study by Tiao and Tsay, the regime for which it was possible to exploit gains was the minority-observations regime (with 15% of the total observations).

5.1.2 MSFEs over different sub-samples

In Table 5 we present the results of the linearity tests applied to different sub-samples. The forecast period is divided in six sub-samples of equal length, each one containing 53 observations. As we can see from the table, there is some suggestions of non-linearities of varying degrees across sub-samples: linearity is strongly rejected in sub-samples 1, 2 and 5, while there appears to be very little evidence of non-linearity for the other sub-samples, especially sub-samples 3 and 4. The forecasts, 49 point forecasts for each horizon (1 to 5), are obtained from the models re-estimated over the appropriate in-sample period, but keeping the same specification as the one selected for the first estimation period 1973.2-1991.6. Since non-linearity is present with different intensities in the various sub-samples, we expect the non-linear models to perform better in periods characterised by strong non-linearities, such as sub-sample 1, 2 and 5. The results in terms of normalised MSFE are presented in Table 6[6]; as above, values greater than one indicate the superiority of the benchmark AR model. The picture obtained from Table 6 is much more informative on the performance of the models than that obtained from the analysis over the whole forecast sample. First of all, from the sub-samples analysis, the DM test reveals significant differences in a larger number of cases, thus making the models more distinguishable. Secondly, even when the linear model seems to outperform the other models over the whole sample, the nonlinear models yield substantial forecasting gains in some of the sub-periods considered. Looking at the results in greater detail, we find that the performance of the GARCH model remains indistinguishable from that of the AR model in all sub-samples, with only one exception (sub-sample 5, h=5) where the GARCH dominates the AR model. This may indicate that the in-mean component of the GARCH specification contributes only marginally to point forecasts. More gains are obtained with the SETAR-2 model, showing substantial improvements over the AR model in the first (h=1,3,5), second (h=3) and fifth (h=1) sub-samples. Finally, SETAR-3 yields more accurate forecasts than the AR model in two cases, namely sub-sample 1 (h=5) and sub-sample 2 (h=3).

Thus, this exercise carried out on a division of the forecast sample by time, has shown that noticeable forecasting gains are provided, especially by the SETAR-2 model, in some of the sub-periods, where the non-linear features of the data appear with greater intensity. This important result questions the often claimed forecasting superiority of the linear models and calls for a more articulate and rigorous analysis of the stochastic characteristics of the period considered.

To complement the above analysis, we proceed to a second exercise, where we divide the sample by the regime of the SETAR

[6] To allow for a straightforward comparison in Table 6 we also report the normalised MSFE for the whole forecasting period (S); the normalised MSFE for the six sub-periods are indicated by S1, S2, S3, S4, S5, S6.

models, to investigate whether the SETAR models record gains relative to the linear models for observations governed by a specific regime.

5.1.3 MSFEs conditional on the regimes of the SETAR models

In this exercise we explore the dependence of forecast performance of the non-linear models on the regime at the forecast origin. In Tables 7A and 7B we report the results for the one-step-ahead point forecasts. The tables report the MSFEs normalised by those for the linear AR model; stars indicate that the forecasting superiority of the models is significant according to the Diebold and Mariano test of equal forecast accuracy. Focusing on the performance of the SETAR models relative to that of the AR counterpart, we can see that significant forecast gains can be achieved in specific regimes with the SETAR models, gains that were not noticeable when the models were assessed unconditionally over the entire forecast period. Interestingly, and in line with previous findings, our results show that gains of the non-linear models over the linear AR alternatives occur, in most cases, for the minority-regime observations. More specifically, there is some evidence of gains for the SETAR-2 model conditional on being in regime 2 (25% observations) over the AR model, while there is no evidence of the SETAR-2 model outperforming the linear counterpart in regime 1, which is the regime with the largest number of observations. Some gains are also obtained for the SETAR-3 model over the AR model when conditioning on regime 3, which is again the regime with fewer observations. Conversely, the AR model does significantly better than the SETAR-3 model in regimes 1 and 2. These results, based on actual data and on a genuine out-of-sample forecast exercise, confirm previous findings by Tiao and Tsay (1994) for the US GNP and those obtained by Clements and Smith (2001) for the exchange rates returns by means of a Monte Carlo study.

Briefly, a look at the performance of the GARCH models versus the AR models shows, again, that there is no much to choose between these two models when they are evaluated on their ability to produce point forecasts. This confirms the marginal contribution that the mean component of the GARCH models plays in forecasting the conditional mean.

5.2. Interval forecasts evaluation

In this section we extend the forecast comparison by evaluating the models on their ability to produce interval forecasts. An interval forecast, or prediction interval, for a variable specifies the probability that the future outcome will fall within a stated interval. The lower and upper limits of the interval forecast are given as the corresponding percentiles. We use central intervals, so that, for example, the 90 per cent prediction interval is formed by the 5^{th} and 95^{th} percentiles. The evaluation of interval forecasts is conducted by means of the likelihood ratio test of correct conditional coverage as recently proposed by Christoffersen (1998). We are interested in detecting whether this comparison reveals gains to the non-linear models, particularly the GARCH models, which were not apparent on MSFE measures.

Christoffersen (1998) shows that a correctly conditionally calibrated interval forecast will provide a hit sequence I_t (for t=1, 2, , T), with value 1 if the realisation is contained in the forecast interval, and 0 otherwise, that is distributed i.i.d. Bernoulli, with the desired success probability *p*. The likelihood ratio test statistic for correct conditional coverage combines a test of unconditional coverage, LR_{UC}, with a test of independence. A sequence of interval forecasts is said to have correct unconditional coverage if $E[I_t]=p$, for all t. Denoting *p* the nominal coverage, n_1 and n_0 the realisations respectively inside and outside the forecast interval, and $\pi=n_1/(n_0+n_1)$ the sample proportion of successes, the test for unconditional coverage is given by:

$$LR_{UC} = -2\log\left(\frac{L(p;.)}{L(p;.)}\right) asy \sim c_1^2$$

The likelihood under the null hypothesis $E[I_t]=p$ is L(p; I_1, I_2, ,I_n)$=(1-p)^{n_0} p^{n_1}$, and under the alternative hypothesis $E[I_t]\neq p$ is L(π; I_1, I_2, ,I_n)$=(1-p)^{n_0} p^{n_1}$. Thus, the test is computed as

$$LR_{UC} = 2\left[n_0 \log(1-p)/(1-p) + n_1 \log(p / p)\right]$$

This test does not have power against the alternative that the zeros and ones are clustered in time-dependent fashion. As stressed by Christoffersen, a simple test for correct unconditional coverage is insufficient in the presence of higher-order moments dynamics (conditional heteroscedasticity, for example). In order to overcome this limitation, Christoffersen proposes a test for independence and a joint test for independence and correct coverage (LR_{IND}).

The LR test for independence assumes a binary first-order Markov chain for the indicator function I_t with transition probability matrix given by:

$$\Pi_1=\begin{bmatrix} 1-p_{01} & p_{01} \\ 1-p_{11} & p_{11} \end{bmatrix}$$

where $\pi_{ij}=\Pr(I_t = j | I_{t-1}=i)$. Under independence $\pi_{ij}=\pi_j$, i,j=0,1 where $\pi_j = \Pr(I_t=j)$. Thus, under the null hypothesis the transition probability matrix is restricted to:

$$\Pi_2=\begin{bmatrix} 1-p_1 & p_1 \\ 1-p_1 & p_1 \end{bmatrix}$$

The π_{ij} and π_i are estimated by their sample frequencies. The unrestricted likelihood for the LR test is given by:

$$L(\Pi_1) = (1-p_{01})^{n_{00}} p_{01}^{n_{01}} (1-p_{11})^{n_{10}} p_{11}^{n_{11}}$$

and the restricted likelihood by:

$$L(\Pi_2) = (1-p_1)^{(n_{00}+n_{10})} p_1^{(n_{01}+n_{11})}$$

where n_{ij} is the number of times event *i* is followed by event *j*. LR_{IND} is asymptotically χ^2 with one degree of freedom under the null hypothesis of independently distributed indicator function values. A combination of these two tests will give a test of correct conditional coverage . The joint test (LR_{CC}) is obtained as the sum of the two LR tests and is asymptotically χ^2 distributed with two degrees of freedom.

In this evaluation analysis of interval forecasts we consider intervals with nominal coverages, p, in the range [0.95-0.05]. The results are presented in Table 8, where for each nominal coverage we report the actual unconditional coverage (π) and the *p*-values of the LR tests presented above. As we can see, for levels of coverages between 95 and 75 per cent the GARCH-M model is the only model to pass all three tests. Both SETAR-2 and SETAR-3 show correct unconditional coverages at these intervals, but the independence test is significant for the SETAR-2 model for all coverages between 90 and 75 per cent, while the SETAR-3 model reject independence only for the 95 and 90 percent intervals. The worse performance is obtained from the AR model which fails the independence test for all intervals from 95 to 75 per cent, as well as the unconditional coverage test for the 75 per cent interval. The good performance of the GARCH model in correspondence of the 95-75 per cent intervals, and of the SETAR-2 models for the 95 per cent interval, imply correct tails forecasts. This result may be seen of special importance in applications with financial variables, where the forecast user is particularly interested in a good performance at the tails of the distribution, that is in correspondence of the big movements .

In Figure 2 we present plots of the interval forecasts obtained from the four competing models, for the 90 per cent coverage. As we can see from the figure, the GARCH model gives interval forecasts that are wider in volatile periods and narrower in tranquil periods. In this case, occurrence of observations outside the intervals are evenly spread over the periods, while in the case of the AR model, and to a lesser extent for the SETAR models, observations outside the intervals are clustered in volatile periods and largely absent from tranquil periods. Thus, a fixed width confident interval as that obtained from AR models is not correctly conditionally calibrated, because it fails to widen when the conditional variance rises and narrow when the conditional variance falls.

Going back to the results in Table 8, we notice that all models produce interval forecasts which are inadequate in the middle range of the distribution (for coverages between 55 and 20 per cent), failing the unconditional coverage tests. We also notice that the adequacy of the interval forecasts improves for the GARCH and the SETAR-3 models at very narrow coverages (15, 10, 5 per cent), while the AR and SETAR-2 models continue to fail the unconditional coverage tests. Thus, there is some suggestion that the forecasts have failed to capture some aspects of the underlying data generating process.[7]

6. Conclusions

In this study we have compared the forecasting performance of alternative univariate time series models for the returns of the Japanese yen exchange rate quoted against the US dollar. The analysis has been carried out on weekly series. Three non-linear models, namely a two-regime SETAR, a three-regime SETAR and a GARCH-M model, have been contrasted with simple linear alternatives (AR processes).

The SETAR and GARCH models have proved successful in describing non-linear features of the data. In particular, the SETAR models have provided strong in-sample evidence for the existence of different regimes, in which the exchange rates exhibit quite different dynamics, while the GARCH models successfully captured the volatility clustering of the returns series.

[7] See Wallis (2002) for a recent discussion of some developments, within the framework of the Pearson s chi-squared goodness-of-fit test, to provide information on the nature of departures from the null hypothesis, with respect to specific characteristics of the distribution of interest such as location, scale and skewness.

The forecast performance of the models has been assessed by means of different forecasting evaluation criteria. First, we have evaluated the models on their ability to produce point forecasts, by comparing MSFEs over the entire forecasting period (1991.6-1997.7). Differences in MSFEs between models were evaluated by means of the Diebold and Mariano test. This analysis did not show significant forecast gains of the non-linear models over the linear benchmark. Then, we have evaluated the models over different out-of-sample periods obtained by splitting the initial forecasting sample in six sub-samples of equal length. This exercise has revealed potential gains especially from SETAR models, in periods characterised by stronger non-linear features. We have also investigated the possibility that non-linear models are more valuable in terms of forecasting accuracy when the process is in a particular regime, specifically, depreciation or appreciation. This analysis has shown that there are significant gains from the SETAR models over the linear AR alternative for the minority-regime observations. Finally, the models have been evaluated on their ability to produce interval forecasts at different nominal coverages. Following Christoffersen (1998), for the evaluation of the interval forecasts we have computed LR tests for unconditional coverage and independence. These tests were then combined into joint tests of conditional coverage. The evaluation of interval forecasts has revealed gains to the non-linear models, particularly the GARCH models, which were not apparent on MSFE measures; moreover this analysis has shown that the static interval forecasts from the AR model are clearly not good conditional interval forecasts. An interesting result is that there was a clear advantage from the GARCH forecasts (and to a lesser extent from the SETAR models) with respect to the AR model at wider coverages (95 to 75 per cent), implying correct coverage of the tails of the distribution. On the other hand, the tests revealed that all models failed to produce forecasts with correct coverage for narrower intervals (between 55 and 20 per cent), suggesting that some aspects of the underlying data generating process have not been adequately captured by the models. However, the forecast performance in the middle range of the distribution of the series may be of minor interest in applications with financial variables, where attention is typically confined to the tails of the distribution (for example, the big movements are the most important for risk management).

To conclude, the empirical analysis conducted in this paper has provided various insights into forecast evaluation. In particular, we have shown that the evaluation results depend not only on the evaluation criterion adopted (i.e. point forecasts versus interval forecasts), but also on the specific sub-samples and levels of coverages considered within each criterion.

Journal of Business and Economic Statistics, 13, 253-263.

Diebold, F.X. and J.A. Nason (1990), Nonparametric exchange rate prediction? *Journal of International Economics*, 28, 315-332.

Engle, R.F. (1982), Autoregressive conditional heteroscedasticity with estimates of the variance of the United Kingdom inflation , *Econometrica*, 50, 987-1008.

Engle, R.F., D.M. Lilien and R.P. Robins (1987), Estimating time varying risk premia in the term structure: the ARCH-M model , *Econometrica*, 55, 391-407.

Glosten, L.R., R. Jagannathan and D. Runkle (1993), On the relation between the expected value and the volatility of the nominal excess return on stocks , *Journal of Finance*, 48, 1779-1802.

Granger, C.W.J. and T. Teräsvirta (1993), *Modelling Nonlinear Economic Relationships*, Oxford University Press, Oxford.

Hsieh, D.A. (1989), A nonlinear stochastic rational expectations model of exchange rates , Unpublished manuscript, Fuqua School of Business, Duke University.

Kräger, H. and P. Kugler (1993), Nonlinearities in foreign exchange markets: a different perspective , *Journal of International Money and Finance*, 12, 195-208.

Luukkonen, R., P. Saikkonen and T. Teräsvirta (1988), Testing linearity in univariate time series models , *Scandinavian Journal of Statistics*, 15, 161-175.

Meese, R. A. and A.K. Rose (1991), An empirical assessment of nonlinearities in models of exchange rate determination , *Review of Economic Studies*, 58, 603-619.

Nelson, D.B (1991), Conditional heteroskedasticity in asset returns: a new approach , *Econometrica* 59, 347-370.

Peel, D.A. and A.E. Speight (1994), Testing for nonlinear dependence in inter-war exchange rates , *Weltwirtschaftliches Archiv*, 130, 391-417.

Priestley, M.B. (1988), *Nonlinear and Non-stationary Time Series Analysis*, Academic Press, London.

Teräsvirta, T. (1994), Specification, estimation and evaluation of smooth transition autoregressive models ; *Journal of the American Statistical Association*, 89, 208-218.

Tiao, G.C. and R.S. tsay (1994), Some advances in non-linear and adaptive modelling in time series , *Journal of Forecasting*, 13, 109-131.

Tong, H. and K.S. Lim (1980), Thresholds autoregression, limit cycles and cyclical data , *Journal of the Royal Statistical Society* B, 42, 245-292.

Tong, H. (1983), *Threshold models in nonlinear time series analysis*, New York, Springer-Verlag.

Wallis, K.F. (2002), Chi-squared Tests of Interval and Density Forecasts, and the Bank of England s Fan Charts , *International Journal of Forecasting* (forthcoming).

Zakoian, J.M (1994), Threshold heteroskedastic models , *Journal of Economic Dynamics and Control*, 18, 931-955.

TABLE 1 DESCRIPTIVE STATISTICS

Japanese yen exchange rate returns	
Mean	-0.000740
Median	0.000000
Maximum	.063120
Minimum	-0.105679
Std. Dev.	0.014186
Skewness	-0.702024
Kurtosis	7.815579
Jarque-Bera	1342.976
Probability	0.000000
Observations	1281

TABLE 2 LINEARITY TESTS - *p*-VALUES

	Entire sample n=1281			Estimation sample n=964			Forecasting sample n=313		
p	*3*	*4*	*5*	*3*	*4*	*5*	*3*	*4*	*5*
RESET *h*=2	0.878	0.712	0.958	0.697	0.522	0.756	0.684	0.644	0.613
RESET *h*=3	**0.025**	**0.098**	**0.036**	0.118	0.165	**0.094**	0.435	0.870	0.464
RESET *h*=4	**0.018**	**0.085**	**0.016**	**0.083**	0.109	**0.043**	0.621	0.877	0.445
S_2, *d*=1	0.141	**0.067**	**0.009**	0.259	0.152	**0.030**	**0.005**	**0.000**	**0.000**
S_2, *d*=2	0.422	0.227	0.172	0.551	0.173	0.141	0.477	0.709	0.789
S_2, *d*=3	**0.017**	**0.037**	0.101	0.155	0.300	0.524	0.317	0.449	0.602
S_2, *d*=4	0.139	**0.071**	**0.087**	**0.056**	**0.020**	**0.024**	0.456	0.556	0.375
S_2, *d*=5	**0.013**	**0.002**	**0.007**	**0.016**	**0.008**	**0.049**	0.374	0.280	**0.080**

p is the autoregressive lag order under the null hypothesis of linearity

TABLE 3 SETAR MODELS SPECIFICATIONS OVER THE ESTIMATION SAMPLE

	SETAR-2		SETAR-3	
	Coeff.	t-value	Coeff.	t-value
$\phi_0^{(1)}$	-0.001	-2.139	-0.003	-3.804
$\phi_1^{(1)}$	0.096	2.698		
$\phi_2^{(1)}$	0.137	3.989		
$\sigma^{(1)}$	0.0134		0.0164	
$T^{(1)}$	736		332	
$\phi_0^{(2)}$	0.001	1.441	0.001	0.283
$\phi_1^{(2)}$	--		0.407	1.708
$\phi_2^{(2)}$	--		0.220	5.176
$\sigma^{(2)}$	0.0150		0.0103	
$T^{(2)}$	222		361	
$\phi_0^{(3)}$	--		0.001	1.160
$\phi_1^{(3)}$	--			
$\phi_2^{(3)}$	--			
$\sigma^{(3)}$	--		0.0140	
$T^{(3)}$	--		265	
$\sigma^{(model)}$	0.0138		0.0137	
d	3		1	
r_1	0.0072		-0.0032	
r_2	--		0.0057	

TABLE 4 FORECASTING PERFORMANCE MSFE AND NORMALISED MSFE

	Number of steps-ahead									
	1		2		3		4		5	
	MSFE	N-MSFE	MSFE	N-MSFE	MSFE	N-MSFE	MSFE	N-MSFE	MSFE	N-MSFE
Naïve	1.8917	--	1.8944	--	1.8915	--	1.8985	--	1.8992	--
Linear AR(2)	1.8929	--	1.8892	--	1.8837	--	1.8948	--	1.8992	--
GARCH-M	1.8980	1.003	1.8903	1.001	1.8785	0.997	**1.8913**	0.998	1.8965	0.999
SETAR-2	**1.8643**	0.985	**1.8777**	0.994	1.8569	0.986	1.9306	1.019**	1.8686	0.984**
SETAR-3	1.9324	1.021	1.9384	1.026**	1.8847	1.001	1.9153	1.011	1.8747	0.987

Note that the value of MSFE has been rescaled by multiplying by 10^4.
The normalised MSFE is calculated as the ratio $MSFE_{NL}/MSFE_{L}$;
*, ** denotes significance of the Diebold-Mariano test at 10% and 5%.

TABLE 5 LINEARITY TESTS BY SUB-SAMPLES- *p*-VALUES

	S1			S2			S3			S4			S5			S6		
p	*3*	*4*	*5*	*3*	*4*	*5*	*3*	*4*	*5*	*3*	*4*	*5*	*3*	*4*	*5*	*3*	*4*	*5*
RESET h=2	**0.008**	**0.070**	0.236	0.971	0.864	0.763	0.169	0.989	0.979	0.437	0.303	0.791	0.698	0.636	0.134	**0.008**	**0.009**	**0.008**
RESET *h*=3	**0.014**	**0.070**	0.287	0.522	0.438	0.463	0.358	0.888	0.872	0.228	0.231	0.109	0.511	0.884	0.322	**0.019**	**0.017**	**0.025**
RESET *h*=4	**0.036**	**0.068**	0.233	0.523	0.348	0.342	0.372	0.746	0.734	0.328	0.387	0.212	**0.027**	0.942	**0.053**	**0.021**	**0.033**	**0.053**
S_2, *d*=1	0.637	0.788	0.666	0.147	0.210	0.207	0.617	0.704	0.314	0.499	0.583	0.450	**0.045**	**0.002**	**0.010**	**0.075**	0.214	0.327
S_2, *d*=2	**0.073**	**0.028**	**0.032**	0.206	0.284	0.213	0.584	0.848	0.718	0.210	0.387	0.232	**0.012**	**0.063**	0.138	0.522	0.625	0.603
S_2, *d*=3	**0.001**	**0.002**	**0.005**	**0.091**	**0.011**	**0.030**	0.835	0.942	0.685	0.781	0.738	0.775	0.576	0.638	0.703	0.712	0.439	0.651
S_2, *d*=4	0.680	0.342	0.281	0.194	**0.095**	0.125	0.484	0.730	0.263	0.355	0.177	0.205	**0.015**	**0.087**	**0.051**	0.269	0.562	0.304
S_2, *d*=5	**0.026**	**0.016**	**0.033**	**0.038**	0.169	0.253	**0.065**	0.102	**0.018**	0.209	0.302	0.112	**0.016**	**0.005**	**0.038**	0.104	0.173	0.334

p is the autoregressive lag order under the null hypothesis of linearity

TABLE 6 NORMALISED MSFE BY SUB-SAMPLES

		Number of steps-ahead				
		1	2	3	4	5
GARCH-M	S	1.003	1.001	0.997	0.998	0.999
	S1	1.017	1.011	1.007	1.007	1.006
	S2	1.022	1.017	1.007	1.009*	1.006
	S3	1.000	1.010	0.999	0.995	1.001
	S4	1.009	1.010	1.009	1.005	1.003
	S5	0.976	0.980	0.999	0.997	0.989**
	S6	0.977	0.985	0.999	1.000	0.995
SETAR-2	S	0.985	0.994	0.986	1.019**	0.984**
	S1	0.957*	0.987	0.937**	1.040*	0.946**
	S2	1.008	1.001	0.939**	1.010	1.024
	S3	1.022	1.047	1.070**	1.000	0.974
	S4	0.983	1.016	1.000	0.983	0.973
	S5	0.929**	0.992	1.012	1.021	0.991
	S6	1.010	1.012	1.012	1.024**	1.019
SETAR-3	S	1.021	1.026**	1.001	1.011	0.987
	S1	1.058	1.021	0.985	1.067**	0.930**
	S2	1.017	1.056	0.889**	0.979	1.013
	S3	1.038	1.007	1.053**	1.021	0.992
	S4	1.032	1.002	0.990	0.995	0.997
	S5	0.982	1.051**	1.039*	1.012	0.987
	S6	0.995	1.034	1.031	1.016	1.006

S refers to the whole forecasting period, S1, S2, S3, S4, S5 and S6 refer to the six subperiods.

*, ** denotes significance of the Diebold-Mariano test at 10% and 5%.

TABLE 7A CONDITIONAL SETAR-2 REGIME 1-STEP-AHEAD FORECASTS

	MSFE			Normalized MSFE		
	Entire sample	Regime 1	Regime 2	Entire sample	Regime 1	Regime 2
T	313	235	78	313	235	78
Linear AR(2)	1.893	1.921	1.791	--	--	--
GARCH-M	1.898	1.939	1.784	1.003	1.009	0.996
SETAR-2	1.864	1.924	1.685*	0.985	1.002	0.941*
ESTAR	1.921	2.002**	1.711*	1.015	1.042**	0.955*

The normalised MSFE is calculated as the ratio $MSFE_{NL}/MSFE_{L}$;
*, ** denotes significance of the Diebold-Mariano test at 10% and 5%.

TABLE 7B CONDITIONAL SETAR-3 REGIME 1-STEP-AHEAD FORECASTS

	MSFE				Normalized MSFE			
	Entire sample	Regime 1	Regime 2	Regime 3	Entire sample	Regime 1	Regime 2	Regime 3
T.	313	116	103	94	313	116	103	94
Linear AR(2)	1.893	1.994	1.816	1.838	--	--	--	--
GARCH-M	1.898	2.059	1.821	1.792	1.003	1.032	1.003	0.975
SETAR-3	1.932	2.093*	1.911*	1.759*	1.021	1.050*	1.052*	0.957*
ESTAR	1.921	2.038	1.914*	1.784	1.015	1.022	1.054*	0.971

The normalised MSFE is calculated as the ratio $MSFE_{NL}/MSFE_{L}$;
*, ** denotes significance of the Diebold-Mariano test at 10% and 5%.

TABLE 8 FORECAST INTERVALS EVALUATION FOR 1-STEP-AHEAD FORECASTS *p*-VALUES

	AR				GARCH				SETAR-2				SETAR-3			
p	π	LR_{UC}	LR_{IND}	LR_{CC}	π	LR_{UC}	LR_{IND}	LR_{CC}	π	LR_{UC}	LR_{IND}	LR_{CC}	π	LR_{UC}	LR_{IND}	LR_{CC}
0.95	*0.939*	0.400	**0.003**	**0.009**	*0.930*	0.120	0.258	0.157	*0.936*	0.278	0.157	0.204	*0.939*	0.400	**0.003**	**0.009**
0.90	*0.901*	0.955	**0.007**	**0.025**	*0.901*	0.955	0.093	0.243	*0.901*	0.955	**0.027**	0.087	*0.891*	0.615	**0.024**	**0.070**
0.85	*0.869*	0.337	**0.011**	**0.025**	*0.853*	0.880	0.166	0.379	*0.853*	0.880	**0.028**	0.088	*0.853*	0.880	0.166	0.379
0.80	*0.837*	0.093	**0.010**	**0.009**	*0.808*	0.712	0.055	0.148	*0.840*	0.067	**0.006**	**0.004**	*0.812*	0.608	0.168	0.339
0.75	*0.802*	**0.030**	**0.009**	**0.003**	*0.773*	0.339	0.066	0.117	*0.789*	0.103	**0.047**	**0.037**	*0.760*	0.670	0.544	0.760
0.70	*0.744*	0.082	0.306	0.130	*0.751*	**0.046**	0.296	0.079	*0.735*	0.174	0.401	0.279	*0.712*	0.629	0.790	0.859
0.65	*0.693*	0.105	0.360	0.176	*0.700*	0.062	0.652	0.159	*0.696*	0.081	0.804	0.212	*0.703*	**0.047**	0.640	0.125
0.60	*0.671*	**0.010**	0.799	**0.034**	*0.645*	0.099	0.538	0.212	*0.665*	**0.018**	0.734	0.059	*0.655*	**0.045**	0.550	0.113
0.55	*0.639*	**0.001**	0.793	**0.006**	*0.620*	**0.012**	0.566	**0.037**	*0.642*	**0.001**	0.767	**0.004**	*0.601*	0.071	**0.031**	**0.019**
0.50	*0.604*	**0.000**	0.947	**0.001**	*0.572*	**0.011**	0.412	**0.028**	*0.578*	**0.006**	0.789	**0.021**	*0.556*	**0.048**	0.069	**0.027**
0.45	*0.540*	**0.001**	0.739	**0.006**	*0.505*	0.052	0.257	0.080	*0.530*	**0.004**	0.693	**0.016**	*0.524*	**0.009**	**0.021**	**0.002**
0.40	*0.492*	**0.001**	0.493	**0.004**	*0.470*	**0.013**	0.450	**0.033**	*0.492*	**0.001**	0.816	**0.004**	*0.476*	**0.006**	0.159	**0.009**
0.35	*0.438*	**0.001**	0.324	**0.004**	*0.415*	**0.017**	0.436	**0.042**	*0.425*	**0.006**	0.950	**0.023**	*0.415*	**0.017**	0.138	**0.019**
0.30	*0.390*	**0.001**	0.825	**0.003**	*0.367*	**0.011**	0.873	**0.038**	*0.377*	**0.004**	0.857	**0.014**	*0.355*	**0.038**	0.436	0.086
0.25	*0.319*	**0.006**	0.396	**0.015**	*0.323*	**0.004**	0.923	**0.015**	*0.316*	**0.008**	0.813	**0.030**	*0.307*	**0.024**	0.159	**0.028**
0.20	*0.249*	**0.034**	0.409	0.076	*0.268*	**0.004**	0.107	**0.004**	*0.259*	**0.012**	0.946	**0.042**	*0.259*	**0.012**	**0.017**	**0.002**
0.15	*0.195*	**0.032**	0.701	0.093	*0.169*	0.346	0.215	0.298	*0.204*	**0.010**	0.693	**0.033**	*0.173*	0.274	**0.006**	**0.013**
0.10	*0.150*	**0.006**	0.214	**0.010**	*0.112*	0.493	0.510	0.636	*0.147*	**0.009**	0.922	**0.033**	*0.105*	0.751	0.370	0.637
0.05	*0.077*	**0.044**	0.903	0.130	*0.054*	0.730	0.175	0.375	*0.077*	**0.044**	0.466	0.101	*0.051*	0.928	0.790	0.961

p indicates the nominal coverage, π indicates the actual unconditional coverage; numbers in bold represent rejections at 5% level of significance

FIGURE 1
EXCHANGE RATE LOG-LEVELS AND RETURNS

1973.1-1997.7

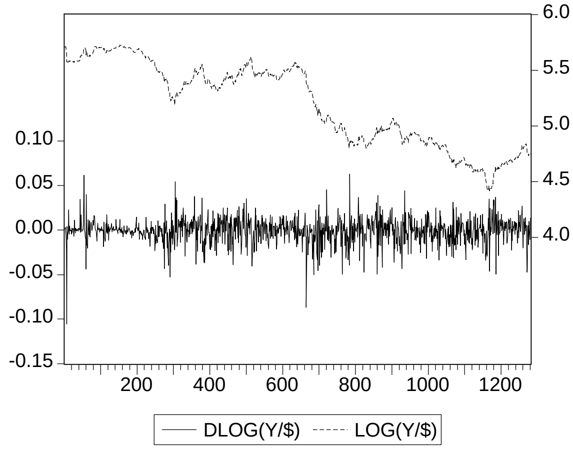
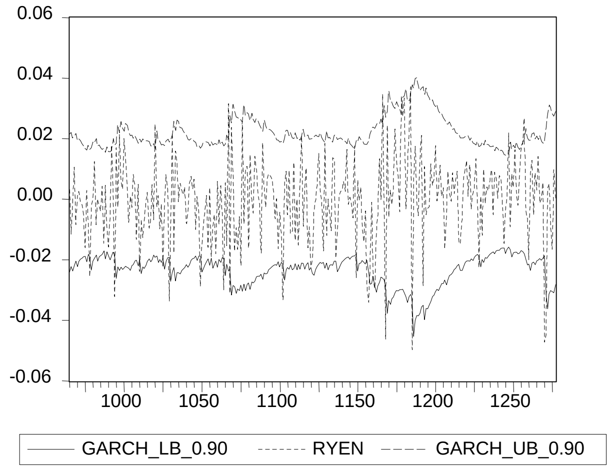


FIGURE 2
90% INTERVAL FORECASTS



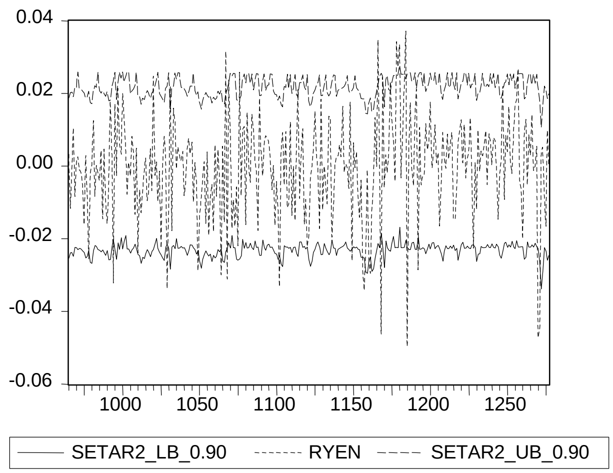


FIGURE 2 CONT.
90% INTERVAL FORECASTS