Available online at www.sciencedirect.com
SCIENCE DIRECT®

ELSEVIER

Journal of Empirical Finance 12 (2005) 490–509

Journal of EMPIRICAL FINANCE

www.elsevier.com/locate/econbase

STAR and ANN models: forecasting performance on the Spanish "Ibex-35" stock index

Jorge V. Pérez-Rodríguez[a,*], Salvador Torra[b], Julián Andrada-Félix[a]

[a]*Department of Quantitative Methods, University of Las Palmas de Gran Canaria, Campus de Tafira, E-35017 Las Palmas, Spain*
[b]*Department of Econometrics, Statistics and Spanish Economy, University of Barcelona, Barcelona, Spain*

Accepted 23 March 2004
Available online 8 December 2004

Abstract

This paper studies whether it is possible to exploit the nonlinear behaviour of daily returns on the Spanish Ibex-35 stock index returns to improve forecasts over short and long horizons. In this sense, we examine the out-of-sample forecast performance of smooth transition autoregression (STAR) models and artificial neural networks (ANNs). We use one-step (obtained by using recursive and nonrecursive regressions) and multi-step-ahead forecasting methods. The forecasts are evaluated with statistical and economic criteria. In terms of statistical criteria, we compared the out-of-sample forecasts using goodness of forecast measures and various testing approaches. The results indicate that ANNs consistently surpass the random walk model and, although the evidence for this is weaker, provide better forecasts than the linear AR model and the STAR models for some forecast horizons and forecasting methods. In terms of the economic criteria, we assess the relative forecast performance in a simple trading strategy including the impact of transaction costs on trading strategy profits. The results indicate a better fit for ANN models, in terms of the mean net return and Sharpe risk-adjusted ratio, by using one-step-ahead forecasts. These results show there is a good chance of obtaining a more accurate fit and forecast of the daily stock index returns by using one-step-ahead predictors and nonlinear models, but that these are inherently complex and present a difficult economic interpretation.

JEL classification: C22; C45; C52
Keywords: Nonlinearities; Statistical criteria; Trading strategies

* Corresponding author. Tel.: +34 928 458222; fax: +34 928 458225/1829.
E-mail addresses: jvperez@dmc.ulpgc.es (J.V. Pérez-Rodríguez), storra@eco.ub.es (S. Torra), jandrada@dmc.ulpgc.es (J. Andrada-Félix).

0927-5398/$ - see front matter
doi:10.1016/j.jempfin.2004.03.001

1. Introduction

The objective of this paper is to investigate whether it is possible to exploit the existence of nonlinearities to improve forecasting over short and long horizons. Therefore, we compare the performance of various nonlinear models in predicting Spanish Ibex-35 stock index returns. The work in this paper is empirical and we do not attempt to explain the results obtained, or those claimed by other researchers, on theoretical grounds.

In many studies of empirical finance, researchers have questioned the hypothesis of Efficient Markets, i.e., that the random walk model is a reasonable description of asset price movement and that linear modelling techniques are appropriate to capture some of the complex models that chartists have observed in the evolution of asset prices and the market negotiation process (e.g., Hinich and Paterson, 1985; White, 1988; Scheinkman and LeBaron, 1989; Hsieh, 1991; Gençay, 1996; Fernández et al., 1999; García and Gençay, 2000). This evidence has provoked theoretical and practical interest in nonlinear financial time-series models, an interest that has increased greatly in recent years because these models are capable of explaining aspects such as the possibility that not all the agents simultaneously receive all the information, the existence of important differences in targets and in negotiation times, or how those agents with more complex algorithms might be able to make better use of the available information.

Once the linear model is rejected in favour of some form of nonlinear model, it is clear that a wealth of possible nonlinear structures can be used to describe and forecast financial and economic time series (e.g., bilinear models, ARCH and its extensions, smooth transition autoregressive models (STAR), artificial neural networks (ANNs), wavelets and even chaotic dynamics). In this paper, we cannot examine all the nonlinear models that might be useful in practice, and therefore we focus on the essentials of a few models which allow various forms of regime switching behaviour, and compare the forecasting performance of different STAR and ANN models in predicting stock index returns.

To study the predictability of stock index returns, we analyse the out-of-sample one-step- and multi-step-ahead forecasts from STAR and ANN models, doing so in two ways. Firstly, we examine whether out-of-sample forecasts generated by nonlinear models are more accurate and preferable to those generated by linear ARMA models for stock index returns, employing statistical criteria such as goodness of forecast measures (i.e., mean absolute error (MAE), mean absolute percentage error (MAPE), root mean squared error (RMSE), Theil's U-statistic), proportion of times the signs of returns are correctly forecasted (Signs) and the directional accuracy test (e.g., Pesaran and Timmermann, 1992), tests for forecast encompassing (e.g. Chong and Hendry, 1986) and equality of accuracy of competing forecasts tests (e.g., Diebold and Mariano's loss differential test (1995)). Secondly, we analyse whether nonlinear ANNs really are superior to linear and STAR models in practice, assessing the relative forecast performance with economic criteria. For example, we use the return forecasts from the different linear and nonlinear models in a simple trading strategy and compare pay-offs to determine whether ANNs are useful forecasting tools for an investor. We also examine the impact of transaction costs on the profits of trading strategies.

This paper is structured as follows: Section 2 describes the empirical methodology, and Section 3, the Ibex-35 data. In Section 4, we examine the predictive capacity of some nonlinear models over a short and a long period, and analyse the trading strategy profits. Finally, Section 5 summarises the most important conclusions.

2. Empirical methodology

In this section, we explain the nonlinear models, forecasting methods and statistical and economic criteria for assessment of the out-of-sample forecasts. Only lagged returns were considered as explanatory variables, because we wanted to analyse the dynamic characteristics of returns from the stock index. Lagged returns are assumed to be necessary in the conditional mean specification for returns, because nonsynchronous trading effects may result in autocorrelation in stock returns.

2.1. Nonlinear models

The nonlinear models we use are regime-switching STAR and ANN models for stock index returns. The STAR and ANN nonlinear models are briefly explained in Sections 2.1.1 and 2.1.2, respectively.

2.1.1. STAR models

These models imply the existence of regimes with potentially different dynamic properties, but with a smooth transition between regimes (e.g., Granger and Teräsvirta, 1993; Teräsvirta et al., 1994). A simple first-order STAR model with two regimes is defined as follows:

$$r_t = \phi_{10} + \sum_{i=1}^{p} \phi_{1i} r_{t-i} + \left[\phi_{20} + \sum_{i=1}^{p} \phi_{2i} r_{t-i} \right] F_{t,d}(s_t; \gamma, c) + \varepsilon_t \tag{1}$$

where r_t are the returns, ϕ_{ij}, (i=1,2, j=0,1,2,. . .,p) are the unknown parameters that correspond to each of the two regimes. $F_{t,d}(s_t;\gamma,c)$ is the transition function, assumed to be twice differentiable and bounded between 0 and 1, γ is the transition rate or smoothness parameter, c is the threshold value which represents the change from one regime to another, and d is the number of lags of the transition variable. This function introduces regime switching and nonlinearity into the parameters of the model. The transition variable, s_t, is usually (but not always) defined as a linear combination of the lagged values of r_t, as: $s_t = \sum_{i=1}^{d} \alpha_i r_{t-i}$. Regarding the choice of transition function, the two most widely used in the literature are the first-order logistic function: $F_{t,d}(s_t;\gamma,c)=\{1+\exp[-\gamma(s_t-c)]\}^{-1}$, $\gamma>0$, in which case the model is called logistic STAR or LSTAR(p;d); and the first-order exponential function, for which: $F_{t,d}(s_t;\gamma,c)=\{1-\exp[-\gamma(s_t-c)^2]\}$, $\gamma>0$, and in this case, the model is called exponential STAR or ESTAR(p;d). In both cases, the transition variable can be any variable in the information set Ψ_{t-1}. Such models are estimated by quasi-maximum likelihood estimations (QMLE). The modelling procedure for building STAR models is carried out in three stages (e.g., Granger and Teräsvirta, 1993, pp. 113–124; Teräsvirta, 1994). Finally, we evaluate the within-sample performance of the estimated STAR models, following Eitrheim and Teräsvirta (1996).

2.1.2. ANN models

ANNs are considered to be a universal approximator in a wide variety of nonlinear patterns, including regime switches and other nonlinearities (e.g., Hornik et al., 1990), and they are good predictors (Swanson and White, 1995, 1997). In this paper, the multilayer perceptron model (MLP) and jump connection nets (JCN) are employed as feedforward networks (Kuan and White, 1994), and we include a partial recurrent network by Elman (1990).

In general, a nonlinear regression model which represents the MLP(p;q) has the following form for a single hidden layer network:

$$r_t = \beta_0 + \sum_{j=1}^{q} \beta_j g\left(\sum_{i=1}^{p} \phi_{ij} r_{t-i} + \phi_{0j} \right) + \varepsilon_t \tag{2}$$

where r_t is the return in t or system output. p is the number of the inputs or lagged stock returns taken as explanatory variables, i.e., $R=(1,r_{t-1},r_{t-2},\cdots,r_{t-p})'$ including a constant term. The parameter vector is $\theta=(\beta',\phi')'$, where $\beta=(\beta_1,\cdots,\beta_q)'$ and $\phi=(\phi_{1j},\cdots,\phi_{pj})'$, $j=1,\ldots,q$, brings together all the network weights, with β_j representing the weights from the hidden to the output unit and ϕ_{ij} the weights from the input layer to the hidden unit j. $g(\cdot)$ determines the connections between nodes of the hidden layer, and is used as the hidden unit activation function to enhance the nonlinearity of the model (i.e., $g(\cdot)$ can take several functional forms, such as the threshold function, which produces binary (± 1) or (0/1) output, or the sigmoid function, which produces an output between 0 and 1). ε_t is a residual *i.i.d*. For an overview and interpretation of Eq. (2), see Kuan and White (1994).

The second model we use is a network with direct connections between the inputs and outputs, called JCN. According to Kuan and White (1994), the parametric specification for the output of the model adds the p-order AR to the MLP network. In this sense, the ANN with a single hidden layer has a linear component augmented by nonlinear terms, and it is written as JCN(p;q) by:

$$r_t = \beta_0 + \sum_{i=1}^{p} \alpha_i r_{t-i} + \sum_{j=1}^{q} \beta_j g\left(\sum_{i=1}^{p} \phi_{ij} r_{t-i} + \phi_{0j} \right) + \varepsilon_t \tag{3}$$

where $\alpha_1,\ldots,\alpha_p$ are direct input–output weights. Eq. (3) nests the linear model because it includes the term $\sum_{i=1}^{p} \alpha_i r_{t-i}$ as a linear autoregressive component.

Finally, the third model used is a partially recurrent network, as proposed by Elman (1990). This has the ability to recognize and, sometimes, to reproduce sequences. This type of ANN is somewhat more complex than the unidirectional ANNs defined by (Eqs. (2) and (3)), has no known linear model characterisation and can be viewed as a nonlinear dynamic latent variable model in econometric terms. In the specific case of a recurrent Elman (p,q) type network, this is characterised by a dynamic structure where the hidden layer output feeds back into the hidden layer with a time delay. In the single hidden layer, this model can take the form:

$$r_t = \beta_0 + \sum_{j=1}^{q} \beta_j z_{j,t} + \varepsilon_t, z_{j,t} = g\left(\sum_{i=1}^{p} \phi_{ij} r_{t-i} + \phi_{0j} + \delta_{ij} z_{j,t-1}\right) \tag{4}$$

where z_j is the output vector of the hidden units, and δ_{ij} are the weights between the hidden units evaluated in t and $t-1$, which introduce a recursive update.

The parameters are estimated by minimising the sum of the squared-error loss: $\min_\theta \sum_{t=1}^{T} [r_t - \hat{r}_t]^2$, where θ is the network weight vector, T is the sample size, and $\hat{r}_t$ is the output value calculated from Eqs. (2)–(4). The most widely used estimation method (the so-called learning rule) of the neural network is error backpropagation (Kuan and White, 1994). For recurrent networks, the model can be estimated by the recurrent backpropagation algorithm and by the recurrent Newton algorithm (e.g., Gençay and Liu, 1997).

2.2. Forecasting methods

We employ one-step- and multi-step-ahead forecasts, because it is useful to compare these models at higher horizons. At this point, two comments should be made: first, although one-step-ahead forecasts are calculated by using the estimated parameters and the actual values for lagged returns, we also estimate the parameters of the models (including linear and nonlinear models) recursively, using ever-larger subsets of the sample data and updating the parameters each period as a new rolling ex ante forecast is constructed. Second, forecasting from nonlinear models beyond one-step-ahead is not straightforward. The multi-step forecast is more difficult than for linear models (the least squares multi-step predictors depend on the known parameters of the models and historical data, but not on the innovation distribution), because exact analytical solutions for a least-squares multi-step forecast are generally not available when the innovation distribution is unknown. Numerical and Monte Carlo simulations or bootstrapping techniques are proposed in the literature to compute multi-step forecasts, when both the nonlinear function and the innovation distribution are known (e.g., Pemberton, 1987; De Gooijer and Bruin, 1998; Clements and Smith, 1999; Lundbergh and Teräsvirta, 2002).

Our simulation-based evaluation of the nonlinear models is different for the STAR and ANN models. We use Monte Carlo evaluation of the forecast performance of the STAR models with resampling from the residuals and we consider 1000 replications.[1] In the case of ANN models, the literature usually considers one-step-ahead prediction, because there

[1] The design of the Monte Carlo experiment for STAR models is as follows. We assume $E[f(\cdot)] \neq \hat{f}(\cdot)$ for forecast periods greater than 2, where $\hat{f}(\cdot)$ is the estimated return from the STAR model by quasi-maximum likelihood. We obtain the realizations of $\{r_t\}_1^T$ for ESTAR and LSTAR. These are generated by considering the parameters to be known and that the errors are distributed as Gaussian white noise with zero mean and unit variances.The h-step forecast is defined as: $\hat{r}_T(h) = \frac{1}{N} \sum_{i=1}^{N} \hat{r}_{T+h}^{(i)}$ where N is the number of replications (i.e., we consider N=1000) and the simulated values $\hat{r}_T(h)$ of $r_T(h)$ are obtained recursively by $\hat{r}_{T+h}^{(i)} = \hat{f}(.)^{(i)} + \varepsilon_{T+h}^{(i)}$, $\varepsilon_{T+h}^{(i)}$ is a random number drawn from a normal distribution with zero mean and unit variance.

is an added difficulty in the application of multi-step prediction. The most common method for neural network multi-step prediction consists of building (training) a predictor for the one-step problem and using it recursively for the corresponding multi-step prediction problem. In this sense, the estimates provided by the model for the next time step are fed back into the input of the model until the desired prediction horizon is reached. Thus, the outputs of the network are fed back to the network as inputs to give a prediction for the next period, and a long-term prediction can thus be produced. We utilize this method, usually called iterated prediction, and we use the same number of data points for the training set as were used for the one-step predictions.[2]

2.3. Statistical and economic criteria for assessment of out-of-sample forecasts

The measures of forecast accuracy and the tests used in this paper are based on h=1,. . .,H prediction periods for stock returns, r_h, called $\hat{r}_h$. For comparison, we use the random walk without drift (RW), linear AR model, STAR and ANN nonlinear models. The statistical measures and tests are briefly explained in Section 2.3.1 and a simple trading strategy is described in Section 2.3.2.

2.3.1. Statistical measures and tests

In order to assess the predictive ability of the different models, we compared the out-of-sample forecasts using two different approaches, because it is generally impossible to specify a forecast evaluation criterion that is universally acceptable. First, we examined the forecast accuracy of all the estimated models by calculating the MAE, MAPE, RMSE, Theil's U-statistic and the proportion of times the signs of returns are correctly forecasted (Signs), as a measure of the market-timing ability of a model. Second, we employed various tests of hypotheses: (i) the Pesaran and Timmermann (DA, 1992) test to examine the directional prediction accuracy of changes, or statistical significance of market-timing ability of a model. Under the null hypothesis, the real and predicted values are independent. (ii) We also tested the forecast encompassing for our out-of-sample forecasts. Following Clements and Hendry (1993), given forecast errors from two models (e_i for model i and e_j from model j), we can test the null hypothesis that one model encompasses the other by running two regressions: the first involves regressing by OLS: $e_{ih}=\alpha_1+\lambda_1(e_{ih}-e_{jh})+u_h$, h=1,. . .,H, thus obtaining the estimated coefficient $\hat{\lambda}_1$. The second involves the regression: $e_{jh}=\alpha_2+\lambda_2(e_{ih}-e_{jh})+u_h$, h=1,. . .,H, and obtaining the estimated coefficient $\hat{\lambda}_2$. In both cases, we analyse whether $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are statistically significant. (iii) We analyse whether the difference between the RMSEs is statistically significant for our out-of-sample forecasts. In this sense, we use the equality of competitive forecasts by the Diebold and Mariano forecast test (1995, DM), which is a fairly general test of the null of equal forecast accuracy for arbitrary loss function (i.e., absolute or squared errors) between two competing models. Given two one-step- or multi-step-ahead predictors, and denoting the corresponding

[2] However, there are other methods to obtain multi-step predictions. For example, Duhoux et al. (2001) suggested a method which consists of training several networks. This method was experimentally found to provide better predictions than the iterated prediction method, but the total number of parameters is proportional to the longest prediction horizon. For an overview, see Boné and Crucianu (2002).

prediction errors by e_{1h} and e_{2h}, then the null hypothesis is tested that $E[d_h]=0$, $d_h=f(e_{1h},e_{2h})$ where d_h is a function of the prediction errors. The above authors assume that the $f(\cdot)$ case is of type $d_h=g(e_{1h})-g(e_{2h})$.

2.3.2. A simple trading strategy

The trading rules considered in this paper are based on a simple market timing strategy, consisting of investing total funds in either the stock market or a risk-free security. The forecast from each predictor is used to classify each trading day into periods "in" (earning the market return) or "out" of the market (earning the risk-free rate of return security). The trading strategy specifies the position to be taken the following day, given the current position and the "buy" or "sell" signals generated by the different predictors. On the one hand, if the current state is "in" (i.e., holding the market) and the share prices are expected to fall on the basis of a sell signal generated by one particular predictor, then shares are sold and the proceeds from the sale invested in the risk-free security (earning the risk-free rate of return r_{fh}). On the other hand, if the current state is "out" and the predictor indicates that share market prices will increase in the near future, the rule returns a "buy" signal and then the risk-free security is sold and shares are bought (earning the market rate of return). Finally, in the other two cases, the current state is preserved (Fernández et al., 2003).

The trading rule return over the predicted period of 1 to H can be calculated as:

$$r=\sum_{h=1}^{H} r_h \cdot I_{bh}+\sum_{h=1}^{H} r_{fh} \cdot I_{sh}+n \cdot \log\left(\frac{1-c}{1+c}\right) \tag{5}$$

where r_h is the market rate of return constructed over the closing price (or level of the Ibex-35 stock index, P_h) on day h; I_{bh} and I_{sh} are indicator variables equal to 1 when the predictor signals are to buy and sell, respectively, and zero otherwise, satisfying the relation $I_{bh}\cdot I_{sh}=0$, $\forall h\in[1,H]$; n is the number of transactions; and c denotes the one-way transaction costs (expressed as a fraction of the price). Regarding the transaction costs, results by Sweeney (1988) suggest that large institutional investors in the mid-1970s could achieve one-way transaction costs in the range of [0.1–0.2%].

In order to assess profitability, it is necessary to compare the return from the trading rule based on the predictors to an appropriate benchmark. To that end, we constructed a weighted average of the return from being long in the market and the return from holding no position in the market and thus earning the risk-free rate of return. The return on this risk-adjusted buy-and-hold strategy can be written as:

$$r_{\mathrm{bha}}=\beta\sum_{h=1}^{H} r_h+(1-\beta)\sum_{h=1}^{H} r_{fh}+2\cdot\log\left(\frac{1-c}{1+c}\right) \tag{6}$$

where β is the proportion of trading days that the rule is in the market.

In this paper, we combine a simple, popular trading strategy known as the filter technique, originally analysed by Alexander (1961) and by Fama and Blume (1966), with parametric and nonparametric forecasts, and compare the return obtained with this risk-adjusted buy-and-hold strategy. In the empirical implementation, the simple rule is modified by introducing a filter to reduce the number of false buy and sell signals, eliminating "whiplash" signals when a selected predictor at date h is close to the closing price at $h-1$. This filtered rule will generate a buy (sell) signal at date h if the predictor is greater than (less than) the closing price at $h-1$ by a percentage δ of the standard deviation σ_{h-1} of the return time series from *1* to $h-1$. Therefore, if $\hat{r}_h$ denotes the prediction for r_h: (i) if $\hat{r}_h > r_{h-1} + \delta \cdot \sigma_{h-1}$ and we are out of the market, a buy signal is generated. If we are in the market, the trading rule suggests we should continue holding the market. (ii) If $\hat{r}_h < r_{h-1} - \delta \cdot \sigma_{h-1}$ and we are in the market, a sell signal is generated. If we are out of the market, we continue holding the risk-free security. A range filter percentage of 0:0.01:1.0 is used because higher filters generally generate no signals.

Given that individuals are generally risk averse, besides the excess return, we also considered a version of the Sharpe ratio (Sharpe, 1966), a measure given by: $RS=\bar{r}/\sigma$, where $\bar{r}$ is the average return of the trading strategy and σ is the standard deviation of daily trading rule returns from adopting the trading strategy in the market. As can be seen, the higher the Sharpe ratio, the higher the mean net return and the lower the volatility returns from being long in the market.

3. Data

We analysed one of the official indexes of the Madrid Stock Market: the Ibex-35 (P_t), an index composed of the 35 most liquid values listed in the Computer Assisted Trading System (CATS), which during the control period had the highest trading volume in cash pesetas. This index is highly representative and is fitted by capitalisation and dividends of the assets included, but not by expansions in capital, and was designed to be used as a reference value in the trade of derivatives products, i.e., options and futures.

This continuous system was introduced onto the Madrid Exchange Market in December 1989. Our study uses the daily closing prices of the Spanish Ibex-35 stock index, from 30 December 1989 to 10 February 2000, with a total of 2520 observations. The series is transformed into logarithms to compute continuous returns, according to the following expression: $r_t = \log\left(\frac{P_t}{P_{t-1}}\right)$, where log is the natural logarithm.

4. Empirical findings on the out-of-sample forecasts

This section focuses on the out-of-sample forecasting ability of the STAR and ANN models in terms of statistical accuracy (Section 4.1) and economic criteria (Section 4.2).[3] The analysis is based on one-step- and multi-step-ahead forecasts, compared with the linear AR and random walk models, considering short and long forecast horizons (h=1,...,H-trading days). Specifically, we chose as 100 the period from

[3] The Brock, Dechert and Scheinkman (BDS, 1987) test, which has a relatively good power to test the induced nonlinearities, rejects the null hypothesis of linearity at 5%, indicating the presence of nonlinearities and, therefore, of complex models in the data. Therefore, the main conclusion is that Ibex-35 stock index returns may be predicted empirically using nonlinear models.

17 July 1999 to 10 February 2000, 200 the period from 3 May 1999 to 10 February 2000, and 500, the period from 12 February 1998 to 10 February 2000, which includes the European stock exchange crisis of 1998. These periods are suitable to describe the robustness of the results achieved.

Let us briefly describe the main characteristics of in-sample nonlinear model estimates: firstly, we use the modelling procedure described by Granger and Teräsvirta (1993) to identify and estimate STAR models. Therefore, we choose p=2 (i.e., 100- and 200-forecast horizons) and p=1 (i.e., 500-trading days) on the basis of SBIC. We also tested whether nonlinear functions of lagged regressor variables contribute significantly to the fit (after correction for a linear AR part), using S_t=r_{t-d}. We considered values for the delay parameter over the range $1 \leq d \leq 12$. On the basis of the linearity test, we choose d=6 when 100- and 200-prediction periods are used, and d=5 for 500-forecast horizons. The parameter in ESTAR and LSTAR is estimated by the QMLE method and the BFGS numerical algorithm.[4] Secondly, we consider that ANN models include one hidden layer[5] and various hidden units or elements of the single hidden layer (q). p, q and $g(\cdot)$ were not chosen a priori. The p-order lagged returns are calculated by sequential validation, and so we estimate ANN models with different values of p and q. The rank of the terms employed is p, q=1,. . .,5. All lagged stock index returns are scaled assuming a uniform distribution within the interval $[-1;1]$. The hidden unit activation function $g(\cdot)$ is the hyperbolic tangent function, which produces a better fit (e.g., Klimasauskas, 1991). All ANN models were trained over a given period (i.e., training sample) using 1500 cycles and cross-validation. The resulting set was used to estimate the neural network weights. To improve on the in-sample fitting performance of the ANN models, the estimated set of weights was used as a set of initial values for training. We used cross-validation strategy in training to avoid overfitting (good in-sample, but poor out-sample performance).[6] The decrease in the error rate in the training and "test" phases was then tested. The output was compared to the sample of original values of the output by comparing the RMSE, which was found to decline as training (i.e., n) progressed. When RMSE reaches a minimum and then starts increasing, this indicates that overfitting may occur. On the basis of the estimated weights from nth training over the training period, out-of-sample forecasts were generated for subsequent "test" periods. In terms of the minimum MSE in the out-of-sample phase, the best adjusted model holds p=2 and q=4 hidden units in the single hidden layer (MLP and JCN) and q=4 (Elman net) for 100- and 200-forecast horizons; and q=1, q=5 (MLP and JCN) and q=4 (Elman net) for 500-prediction periods.

Once each model has been estimated, we construct its out-of-sample forecast. The empirical findings are discussed in the following sections.

[4] We did not include GARCH models in the set of forecasting models because these models parameterise the conditional variance, whereas the object to be forecasted is the stock index return, not its volatility.

[5] As there is no reliable method of specifying the optimal number of hidden layers, we specified one hidden layer, because this is generally effective in capturing nonlinear structures (e.g., Adya and Collopy, 1998).

[6] For example, if we consider H=200, the training phase of the ANN was performed with 1855 observations, whereas in the "test" phase 463 observations were used, both sizes being randomly determined. The 200 final observations were set aside to make predictions and were not in any way used in the estimation.

4.1. Statistical assessment of the out-of-sample forecast

In this section, we discuss the model selection procedures described in Section 2.3.1., focussing on the relative merits of ANNs and its competitors. We first examine the goodness of forecast measures and test the market-timing ability of the models, and then consider the full set of results at 100-, 200- and 500-forecast horizons and the forecasting methods commented in Section 2.2 (i.e., one-step-ahead predictions, one-step-ahead forecast in recursive regressions, and multi-step forecasts). These are shown in each table as Panels A, B and C, respectively. Finally, we use a similar procedure to comment on the results of the forecast encompassing test and the forecast accuracy test. These results are as follows:

(i) Table 1 gives the results of the goodness of forecast, Signs and P-value of the Pesaran and Timmermann test for all models. The main conclusions are as follows: first, the results for forecast horizons and forecasting methods are heterogeneous in terms of the MAE, MAPE and RMSE measures. However, the ANN forecast is a slight improvement on the linear AR and STAR models, because in many cases ANNs have lower forecast errors (according to the above measures). Nevertheless, taking into account the number of times where ANNs are better than the linear AR and STAR models, the best forecast horizon for our ANN technique is 100-forecast periods and the best forecasting method is the one-step-ahead forecasting method. However, other results are less favourable for the ANN models. Their performance deteriorates when the forecast horizon is extended, and the ANN and STAR models are inferior to the linear AR model, in terms of the one-step-ahead forecast when the recursive regression technique is used (Panel B), and also in terms of MAE results for all forecast horizons. This fact implies that AR is capable of successfully describing the evolution of the Ibex-35 stock index returns.

Second, it is clear that ANN models surpass the nondrift random walk method by using Theil's U-statistic at any horizon and forecasting method. Thus, the Ibex-35 stock index return can be predicted and the random walk model does not provide a reasonable description of asset price movement.

Third, ANN also seems to provide a slightly improved forecast with respect to linear AR and STAR models, according to the market-timing ability (Signs). The signs of returns are correctly forecasted 55–57% in the case of one-step-ahead forecasts (recursive and nonrecursive regressions)[7]. However, we do not reject the null hypothesis of DA test that forecasts and realizations are independent at any significance level. Therefore, we consider the market-timing ability of a model is not statistically significant at any level, except in the case of one-step forecasts obtained by recursive regressions for Elman and MLP at 100- and 200-one-step forecast horizons, and at the 10% significance level. This indicates that neural networks do not perform much better than linear and STAR models, because we cannot reject the null hypothesis that real and predicted values are independent in one-step-ahead forecasts. In summary, the ANN models give best results, with a slightly superior out-of-sample performance to that of the other models (i.e., Theil's U-statistic in

[7] In the case of multi-step forecasts, Signs are equal for all models because the forecasts take positive values and tend to be constant (i.e., in the linear AR model they tend to the mean of the process). Thus, Panel C does not show the DA test.

Table 1
Goodness of forecast and DA test for all models (including random walk model)

Models	H=100						H=200						H=500					
	MAE	MAPE	RMSE	U-Theil	Signs	DA	MAE	MAPE	RMSE	U-Theil	Signs	DA	MAE	MAPE	RMSE	U-Theil	Signs	DA
Panel A: One-step-ahead																		
RW	–	–	–	1.0000	–	–	–	–	–	1.0000	–	–	–	–	–	1.0000	–	–
AR	0.9828	101.47	1.2040	0.8717	0.54	0.93	0.9042	112.77[a]	1.1353	0.8821	0.505	0.84	1.2325	122.60[a]	1.6758	0.8956	0.53	0.34
ESTAR	0.9879	104.12	1.2049	0.8585	0.52	0.73	0.9104	116.31	1.1417	0.8782	0.515	0.70	1.2314[a]	121.93	1.6750[a]	0.8908	0.544	0.20
LSTAR	0.9837	104.56	1.2065	0.8515	0.56[a]	0.60	0.9049	115.22	1.1345	0.8750	0.515	0.98	1.2378	137.38	1.7009	0.8759	0.486	0.42
MLP	0.9795[a]	100.40[a]	1.2030[a]	0.8810	0.56[a]	0.92	0.9000[a]	115.99	1.1309[a]	0.8883	0.555	0.44	1.2356	145.04	1.6810	0.9229	0.486	0.61
JCN	1.0951	102.44	1.3294	0.8068[a]	0.41	0.72	0.9014	116.90	1.1331	0.8779	0.534	0.99	1.2465	213.46	1.6835	0.8814	0.464	0.34
Elman	1.0185	114.15	1.2348	0.8561	0.52	0.48	0.9095	118.93	1.1410	0.8675[a]	0.525	0.83	1.2586	243.30	1.6838	0.8076[a]	0.55[a]	0.50
Panel B: One-step-ahead (recursive regressions)																		
RW	–	–	–	1.0000	–	–	–	–	–	1.0000	–	–	–	–	–	1.0000	–	–
AR	0.9821[a]	101.45	1.2029	0.8708	0.53	0.79	0.9031[a]	112.19[a]	1.1342[a]	0.8844	0.50	0.86	1.1344[a]	147.55	1.6757[a]	0.8965	0.517	0.65
ESTAR	0.9875	104.84	1.2049	0.8653	0.54	1.00	0.9137	115.34	1.1434	0.8780	0.50	0.58	1.2365	135.21	1.6796	0.8837	0.524	0.71
LSTAR	0.9831	104.16	1.1965[a]	0.8515	0.56	0.79	0.9044	115.14	1.1344	0.8714	0.51[a]	0.59	1.2358	134.09[a]	1.6936	0.8865	0.546[a]	0.17
MLP	1.0207	100.11[a]	1.2666	0.8935	0.57[a]	0.32	0.9118	119.55	1.1553	0.8811	0.51[a]	0.07[a]	1.2393	179.85	1.6896	0.8987	0.518	0.67
JCN	0.9874	102.45	1.2070	0.8809	0.54	0.71	0.9208	115.03	1.1487	0.8921	0.50	0.23	1.2490	185.39	1.6875	0.8720	0.508	0.80
Elman	1.1849	121.18	1.4483	0.7933[a]	0.54	0.08[a]	1.1201	118.51	1.4048	0.7308[a]	0.50	0.44	1.1498	251.21	1.9413	0.7225[a]	0.470	0.64
Panel C: Multi-step-ahead																		
RW	–	–	–	1.0000	–	–	–	–	–	1.0000	–	–	–	–	–	1.0000	–	–
AR	1.0000	100.17	1.2237	0.9517	0.61	–	0.9034	105.43	1.1347[a]	0.9507	0.555	–	1.2305	138.51	1.6870[a]	0.9679	0.546	–
ESTAR	0.9995	98.033	1.2238	0.9551	0.61	–	0.9036	104.42	1.1352	0.9556	0.555	–	1.2306	131.66[a]	1.6871	0.9716	0.548	–
LSTAR	0.9967	97.377[a]	1.2221	0.9449	0.61	–	0.9036	104.59	1.1351	0.9548	0.555	–	1.2306	131.66[a]	1.6871	0.9727	0.548	–
MLP	0.9697[a]	99.169	1.2136	0.8290	0.61	–	0.9061	100.24[a]	1.1368	0.9819	0.555	–	1.2285[a]	190.29	1.6875	0.9343	0.548	–
JCN	0.9716	102.54	1.2120[a]	0.8282[a]	0.61	–	0.9033[a]	106.42	1.1351	0.9467	0.555	–	1.2317	153.88	1.6882	0.9576	0.548	–
Elman	0.9803	100.36	1.2142	0.8676	0.61	–	0.9624	167.44	1.1904	0.8473[a]	0.555	–	1.4966	197.41	1.9195	0.7603[a]	0.548	–

The table reports the goodness of forecast and DA test of alternative forecasting models. P-values appear for DA test.

[a] Indicates the best result between models in the same column.

Elman net and Signs in MLP) for all forecasting methods and short and long forecast horizons.

(ii) Table 2 reports the results of Chong and Hendry's forecast encompassing test, which makes a two-by-two comparison of the seven competitors (including a random walk model). This table shows the *P*-values associated with the heteroskedasticity-robust *t*-statistics on the null hypothesis λ_1=0 and λ_2=0, respectively, for all *i*–*j* comparisons (*i*,*j*=RW, AR, ESTAR, LSTAR, MLP, JCN and Elman). For each forecasting method, we have three matrices, for which the following aspects are considered: let b_{ij} (*i*,*j*=RW, AR, ESTAR, LSTAR, MLP, JCN and Elman) be the general element for each matrix, where *i* is the *i*th row and *j* is the *j*th column. So, for example, when referring to $b_{\mathrm{ESTAR,RW}}$ we consider the *P*-value for $\hat{\lambda}_1$ (lower triangular matrix), and when we refer to $b_{\mathrm{RW,ESTAR}}$ we consider the *P*-value of $\hat{\lambda}_2$ (upper triangular matrix).

We observe two general results from Table 2. Firstly, the forecast encompassing test shows that the random walk model always performs worse than the AR, STAR and ANN models. The predictions of the AR, STAR and ANN models are conditionally more efficient than those of the random walk model without drift, in all forecast horizons and for all forecasting methods considered. Given that $b_{i,\mathrm{RW}}$>0.10 (*i*=AR, ESTAR, LSTAR, MLP, JCN, Elman) indicates that $\hat{\lambda}_1$ is not statistically significant (i.e., *P*-value higher than 10% in the lower triangular matrix) and $b_{\mathrm{RW},j}$>0.10 (*j*=AR, ESTAR, LSTAR, MLP, JCN, Elman) indicates that $\hat{\lambda}_2$ is statistically significant (i.e., *P*-value lower than 10% in the upper triangular matrix), we reject the null hypothesis in favour of the alternative hypothesis that the AR, STAR and ANN models encompass the random walk model at the 10% significance level. Secondly, we cannot consider that ANN models surpass the AR and STAR models in all cases. We hardly ever reject the possibility that neither model encompasses the other for the remaining models and for the different prediction methods, because neither $\hat{\lambda}_1$ nor $\hat{\lambda}_2$ are significant (b_{ij}>0.10, $\forall i \neq j$). Furthermore, this is always true for the multi-step-ahead prediction method (Panel C). However, another aspect stands out (one that is also evident at 100- and 200-day forecast horizons): a small number of results for the one-step-ahead prediction model (in both the recursive and the nonrecursive cases) show that the MLP and JCN networks are slightly better (i.e., conditionally more efficient) than the linear AR and STAR models. For example, in Panel A and for the 200-day horizon, if we consider the ESTAR row and the AR column (*P*-value 0.18) and the AR row and the ESTAR column (*P*-value 0.06), then the null hypothesis is rejected in favour of the alternative that ESTAR encompasses AR at the 10% significance level. Furthermore, MLP encompasses AR at the 10% significance level (*P*-values 0.91 and 0.10, respectively). Finally, if we consider the MLP row and the ESTAR column (*P*-value 0.60) and the ESTAR row and the MLP column (*P*-value 0.05), we conclude that MLP encompasses ESTAR at the 5% and 10% levels. Therefore, we conclude that ANN one-step-ahead forecasts are slightly to be preferred to other forecast combining techniques on the basis of encompassing tests, and that MLP is a good ANN. Similar results can be observed in Panel B for the MLP model. However, the AR, ESTAR, LSTAR, MLP and JCN models encompass Elman net at any significance level.

(iii) Finally, Table 3 shows the Diebold and Mariano forecast accuracy test for the MAE loss function. This test examines whether the difference in the MAE of the forecasts

Table 2

P-values associated with the heteroskedasticity-robust t-statistics on λ_1 and λ_2 for all possible i–j comparisons based on OLS regressions

Models	H=100							H=200							H=500						
	RW	AR	ESTAR	LSTAR	MLP	JCN	Elman	RW	AR	ESTAR	LSTAR	MLP	JCN	Elman	RW	AR	ESTAR	LSTAR	MLP	JCN	Elman
Panel A: One-step-ahead																					
RW	–	0.00	0.00	0.00	0.00	0.00	0.00	–	0.00	0.00	0.00	0.00	0.00	0.00	–	0.00	0.00	0.00	0.00	0.00	0.00
AR	0.86	–	0.04	0.28	0.05	0.46	0.41	0.35	–	0.06	0.61	0.10	0.34	0.69	0.83	–	0.55	0.50	0.46	0.42	0.50
ESTAR	0.99	0.78	–	0.77	0.04	0.61	0.50	0.45	0.18	–	0.14	0.05	0.09	0.49	0.88	0.89	–	0.47	0.54	0.66	0.55
LSTAR	0.95	0.62	0.61	–	0.21	0.57	0.36	0.40	0.91	0.48	–	0.36	0.37	0.88	0.49	0.11	0.10	–	0.11	0.11	0.14
MLP	0.65	0.73	0.34	0.28	–	0.51	0.18	0.94	0.91	0.60	0.82	–	0.92	0.91	0.51	0.22	0.19	0.57	–	0.21	0.25
JCN	0.18	0.42	0.30	0.26	0.51	–	0.72	0.49	0.45	0.70	0.42	0.46	–	0.46	0.97	0.61	0.94	0.51	0.64	–	0.68
Elman	0.24	0.03	0.02	0.02	0.01	0.04	–	0.24	0.21	0.82	0.19	0.09	0.09	–	0.05	0.14	0.13	0.76	0.15	0.13	–
Panel B: One-step-ahead (recursive regressions)																					
RW	–	0.00	0.00	0.00	0.00	0.00	0.00	–	0.00	0.00	0.00	0.00	0.00	0.00	–	0.00	0.00	0.00	0.00	0.00	0.00
AR	0.86	–	0.03	0.28	0.10	0.73	0.15	0.38	–	0.03	0.75	0.10	0.25	0.32	0.89	–	0.50	0.24	0.37	0.36	0.86
ESTAR	0.87	0.58	–	0.20	0.07	0.98	0.18	0.42	0.20	–	0.04	0.09	0.14	0.46	0.84	0.08	–	0.52	0.88	0.90	0.63
LSTAR	0.95	0.62	0.62	–	0.02	0.57	0.14	0.32	0.72	0.42	–	0.87	0.36	0.29	0.66	0.05	0.10	–	0.20	0.33	0.40
MLP	0.80	0.80	0.15	0.02	–	0.02	0.99	0.46	0.16	0.13	0.07	–	0.95	0.73	0.24	0.00	0.02	0.48	–	0.22	0.46
JCN	0.65	0.30	0.38	0.16	0.02	–	0.17	0.93	0.22	0.18	0.02	0.55	–	0.76	0.83	0.01	0.07	0.77	0.43	–	0.65
Elman	0.11	0.00	0.00	0.00	0.00	0.00	–	0.12	0.00	0.00	0.00	0.00	0.00	–	0.13	0.00	0.00	0.00	0.00	0.00	–
Panel C: Multi-step-ahead																					
RW	–	0.00	0.00	0.00	0.00	0.00	0.00	–	0.00	0.00	0.00	0.00	0.00	0.00	–	0.00	0.00	0.00	0.00	0.00	0.00
AR	0.27	–	0.24	0.11	0.81	0.40	0.92	0.61	–	0.05	0.16	0.00	0.00	0.27	0.06	–	0.00	0.00	0.70	0.00	0.00
ESTAR	0.27	0.50	–	0.38	0.84	0.48	0.14	0.61	0.05	–	0.62	0.27	0.00	0.35	0.06	0.00	–	0.80	0.01	0.00	0.00
LSTAR	0.27	0.29	0.42	–	0.69	0.54	0.27	0.61	0.15	0.63	–	0.01	0.00	0.36	0.06	0.00	0.82	–	0.01	0.01	0.00
MLP	0.26	0.68	0.61	0.52	–	0.45	0.74	0.61	0.00	0.25	0.01	–	0.11	0.54	0.06	0.64	0.01	0.01	–	0.00	0.60
JCN	0.28	0.64	0.69	0.73	0.77	–	0.63	0.61	0.00	0.00	0.00	0.18	–	0.78	0.06	0.00	0.00	0.00	0.00	–	0.00
Elman	0.26	0.84	0.11	0.19	0.83	0.41	–	0.61	0.11	0.28	0.28	0.45	0.73	–	0.06	0.00	0.00	0.00	0.26	0.00	–

For each horizon and panel, we have a block with two triangular matrices. Upper triangular matrix corresponds to P-values for $\hat{\lambda}_2$ and lower triangular matrix corresponds to P-values for $\hat{\lambda}_1$. In this table, when we refer to P-value for $\hat{\lambda}_1$, the row is the ith model and the column is the jth model. When we refer to P-value for $\hat{\lambda}_2$, the row is the jth model and the column is the ith model.

Table 3
Diebold and Mariano test for $d_h = |e_{1h}| - |e_{2h}|$ (MAE, upper triangular matrix in each block of the panel)

e_1	H=100							H=200							H=500						
e_2	RW	AR	ESTAR	LSTAR	MLP	JCN	Elman	RW	AR	ESTAR	LSTAR	MLP	JCN	Elman	RW	AR	ESTAR	LSTAR	MLP	JCN	Elman
Panel A: One-step-ahead																					
RW	–	5.1	5.0	5.0	4.9	5.1	5.4	–	7.6	7.5	7.3	6.9	6.5	7.7	–	13.9	13.8	13.3	13.4	13.5	15.0
AR		–	−0.5	−0.1	0.7	1.3	0.2		–	−1.7	−1.4	0.5	0.2	−1.3		–	0.6	−0.6	−0.9	−2.2	−2.9
ESTAR			–	0.4	0.6	1.0	0.4			–	−0.2	1.2	0.7	0.1			–	−0.8	−1.1	−2.4	−3.1
LSTAR				–	0.4	0.8	0.2				–	1.3	0.8	0.2				–	0.2	−0.8	−1.6
MLP					–	1.5	−0.1					–	−0.3	−0.9					–	−2.2	−2.2
JCN						–	−0.7						–	−0.6						–	−1.4
Elman							–							–							–
Panel B: One-step-ahead (recursive regressions)																					
RW	–	5.1	5.0	5.0	4.2	4.7	4.2	–	7.6	7.3	7.6	6.6	6.7	1.8	–	13.9	3.8	13.7	12.4	13.1	4.3
AR		–	−1.1	−0.1	−1.8	−0.6	−3.4		–	−2.3	−0.2	−0.6	−2.1	−3.8		–	−0.4	−0.2	−0.6	−2.2	−7.2
ESTAR			–	0.5	−1.5	0.0	−3.3			–	1.6	0.1	−0.8	−3.6			–	0.1	−0.3	−1.8	−6.8
LSTAR				–	−1.6	−0.3	−3.5				–	−0.5	−1.7	−3.8				–	−0.4	−1.7	−6.9
MLP					–	1.7	−2.8					–	−0.6	−3.7					–	−1.2	−6.6
JNC						–	−3.5						–	−3.4						–	−6.6
Elman							–							–							–
Panel C: Multi-step-ahead																					
RW	–	4.5	4.5	4.5	5.1	5.1	4.9	–	6.8	6.8	6.8	6.8	6.8	5.5	–	12.5	12.5	12.5	12.5	12.5	4.2
AR		–	0.7	2.2	1.8	1.5	1.7		–	−0.9	−0.9	−1.1	−0.1	−2.6		–	−0.1	−0.2	0.7	−0.6	−7.2
ESTAR			–	2.2	1.7	1.5	1.6			–	0.7	−1.1	0.4	−2.6			–	−0.6	0.6	−0.6	−7.2
LSTAR				–	1.6	1.4	1.5				–	−1.1	0.4	−2.6				–	0.6	−0.5	−7.3
MLP					–	−1.1	−1.7					–	0.9	−2.8					–	−1.2	−6.8
JCN						–	−1.2						–	−2.6						–	−7.1
Elman							–							–							–

Critical distribution values for $N(0,1)$ are 1.645, 1.96, 2.576 at 10%, 5% and 1%, respectively.

obtained from two models is statistically significant. The DM test based on the difference between two MAE, $d_h=|e_{1h}|-|e_{2h}|$, is shown in the upper triangular matrix of each forecast horizon. We have omitted the lower triangular matrix because the original matrix is diagonal symmetric apart from multiplication by -1. The e_2 forecast errors are shown in columns and the e_1 errors in rows.[8] Two aspects are particularly noteworthy: firstly, the DM test is conclusive when the random walk model is compared with the other models for all horizon and forecasting methods. If we use the random walk as e_1 and use different forecast errors (i.e., AR, ESTAR, LSTAR, MLP, JCN and Elman) as e_2, the DM test is positive and statistically significant at the 5% significance level in all cases. The model selected as 2 always has a smaller MAE than the random walk model. The latter performs worse than the AR, STAR and ANN models and, therefore, both linear and nonlinear models improve on random-walk forecasts. Secondly, if we compare the remaining models, the results are not conclusive for all forecast horizons and prediction methods; thus, the difference between the two functions of true forecast errors is identically zero.[9] Furthermore, the fact that the null hypothesis is not rejected might be because the models being compared are nested models (except Elman net), and so this test is not consistent against generic nonlinear alternatives, as pointed out by Corradi and Swanson (2002).[10]

In summary, the statistical criteria results constitute weak evidence in favour of the existence of nonlinearities that are exploitable to improve forecasts. ANN models only perform better than AR and STAR models in terms of goodness of forecasts and in some cases of the forecast encompassing test. Linear AR and nonlinear models surpass RW in all statistics and tests. However, in this aspect, if there is no rule about profits, and no profits are made over a long period, then the weak hypothesis of Efficient Markets should not be rejected (see, e.g., Granger, 1992).

4.2. Assessing relative forecast performance with economic criteria in a simple trading strategy

In this section, we assess the economic criteria, using the return forecasts (one-step and multi-step forecasts) from the different models in a simple trading strategy and comparing

[8] It should be noted that when a finite sample (i.e., H=100) is utilised, Harvey et al. (1998, 1999) recommend modifying the DM test. They prefer to compare the modified Diebold and Mariano statistic with critical values from the t-Student distribution with $H-1$ degrees of freedom, rather than use the normal standard.

[9] For example, for 200-day forecast horizons and for the remaining models, Panel A shows that the null hypothesis that two forecasts have an equal MAE is not rejected at the 5% significance level, except the AR row and the ESTAR column at 10% ($DM=-1.7$) where AR is significantly smaller than that of the ESTAR forecasts. For Panel B, we reject the null hypothesis in the AR row and the ESTAR column and in AR-JCN at 5%. As DM<0, AR is preferred to the ESTAR and JCN models. We also reject the null hypothesis for all models against the Elman model at 5%. Finally, for Panel C, we reject the null hypothesis in Elman net against other models (except the random walk model).

[10] It should also be noted that parameter estimation error may be important, but is ignored in our tests (see, e.g., West, 1996; Corradi and Swanson, 2001, 2002); also noteworthy is the issue of joint comparison of more than two competing models (e.g., White, 2000). In particular, the Diebold and Mariano test for nested models may be invalid. Corradi and Swanson (2002) propose a variant of the Bierens ICM test that captures both the time series structure of the data as well as the contribution of the parameter estimation error, which is consistent against generic nonlinear alternatives, and which is designed for comparing nested models.

the pay-offs to determine whether ANNs are useful forecasting tools for an investor (Leitch and Tanner, 1991; Satchell and Timmermann, 1995).[11]

In order to assess the economic significance of predictable patterns in the Ibex-35 series, it is necessary to consider explicitly how investors may exploit the computed local predictions as trading rules. Table 4 compares forecasts for 100-, 200- and 500-forecast horizons (excluding random walk) using economic criteria. We use the trading rule return and the return on risk-adjusted buy-and-hold strategy, assuming a transaction cost of 0.20% and also consider the modified version of the Sharpe risk-adjusted ratio. We analyse a transaction cost of 0.20% because the mean intermediation commission in Spain is similar to this percentage of the quantity invested.[12] The results are averaged with respect to filter values for which the trading strategy produces different economic outcomes. In every case, we used the filter technique presented in Section 2.3.2, considering the different forecasting methods (Panels A, B and C, respectively).

Looking first at the trading rule return (r), we compare the return from the trading rule based on the predictors, with the return on the risk-adjusted buy-and-hold strategy being taken as a benchmark. The model selection procedure is as follows: identify the best value of r between all models and then compare r with its r_{bha}, taking into account that a good model has $r>r_{\mathrm{bha}}$. Our main conclusions are, firstly, that nonnegative mean net returns are generally obtained for short and long forecast horizons and for all forecasting methods. Secondly, given that neural networks were not found to perform much better than linear or STAR models in terms of DA, we would not find it surprising if ANNs did not provide significantly higher trading profits. However, nonlinear models generally yield marginally significant, slightly higher trading profits for all forecast horizons and for the two one-step-ahead prediction methods. STAR and ANNs yield higher mean net returns than does the linear model in one-step-ahead forecasts (nonrecursive and recursive). However, except for the MLP model in one-step-ahead nonrecursive forecasts, we cannot guarantee that $r>r_{\mathrm{bha}}$, for all models in all forecasting periods. For example, JCN performs best with the 100-forecast horizon, while MLP performs best with the 200- and 500-forecast horizons. However, ESTAR performs best with the 100- and 500-forecast horizons and MLP for the 200-forecast period in terms of one-step-ahead forecasts obtained by recursive regressions. Thirdly, we found no significant differences between the forecasting models when multi-step-ahead forecasts were used.

Finally, Table 4 provides the mean Sharpe risk-adjusted ratio derived from a simple trading strategy. In this case, the findings are similar to those for the mean net returns. Thus, the mean Sharpe ratio is nonnegative and nonlinear models generally yield higher

[11] As shown by Leitch and Tanner (1991) and by Satchell and Timmermann (1995), the use of statistical or economic criteria can lead to very different outcomes. Satchell and Timmermann (1995) have pointed out that standard forecasting criteria are not necessarily particularly well suited to assess the economic value of predictions of a nonlinear process.

[12] These costs also include the operating fees charged by the Stock Exchange (proportional to the value of the operation), the commission for management of stocks and shares (both for custody and for payment operations) and the (usually nominal) commission charged by the Stock Liquidation and Compensation Service. Initially, we used one-way transaction costs of 0.15%, 0.20% and 0.25%, because we also wished to investigate the robustness of the results for different costs. Because the results are similar, we have omitted those for 0.15% and 0.25%. Complete results are available upon request to the authors.

Table 4
Mean net return and mean Sharpe ratio over different horizons of prediction

	Mean net return						Mean Sharpe ratio					
	H=100		*H*=200		*H*=500		*H*=100		*H*=200		*H*=500	
	r	r_{bha}	r	r_{bha}	r	r_{bha}	r	r_{bha}	r	r_{bha}	r	r_{bha}
Panel A: One-step-ahead												
AR	0.00144	0.00138	0.00057	0.00068	0.00056	0.00053	0.14615	0.18989	0.06044	0.0834	0.04161	0.1017
ESTAR	0.00137	0.00162	0.00054	0.00064	0.00053	0.00051	0.13265	0.18286	0.0595	0.08758	0.0439	0.05771
LSTAR	0.00113	0.00094	0.00034	0.00067	0.00063	0.00061	0.15019	0.27513	0.03456	0.08351	0.04583	0.04817
MLP	0.00142	0.00132	0.00079[a]	0.00071	0.00064[a]	0.00055	0.14304	0.19473	0.08059	0.09007	0.05049[a]	0.04992
JCN	0.00172[a]	0.00157	0.00071	0.00065	0.00001	0.00019	0.16024	0.18949	0.07778	0.09645	0.00035	0.22058
Elman	0.00161	0.00159	0.00042	0.00076	0.00049	0.00039	0.16012	0.2275	0.03904	0.08148	0.04448	0.06898
Panel B: One-step-ahead (recursive regressions)												
AR	0.00135	0.00133	0.00058	0.00069	0.00047	0.00044	0.13934	0.19154	0.06111	0.08306	0.03792	0.08863
ESTAR	0.00158[a]	0.00152	0.00049	0.00069	0.00059[a]	0.00051	0.15646	0.1832	0.04835	0.0829	0.05178	0.05862
LSTAR	0.00106	0.00089	0.00025	0.00045	0.00039	0.00036	0.14463	0.28498	0.061	0.26132	0.03822	0.07297
MLP	0.00136	0.00149	0.00106[a]	0.00072	0.00038	0.00065	0.12282	0.18821	0.10459[a]	0.08246	0.02575	0.04752
JCN	0.00115	0.0013	0.00057	0.00079	0.00047	0.0005	0.15793	0.40076	0.05182	0.08069	0.04151	0.0599
Elman	−0.00003	0.00057	−0.00058	0.00055	−0.00006	0.00024	0.00998	0.39572	−0.07114	0.08708	−0.01356	0.08083
Panel C: Multi-step-ahead												
AR	0.00201	0.00214	0.00088	0.00089	0.00076	0.00076	0.16478	0.17736	0.07742	0.07888	0.04523	0.04551
ESTAR	0.00203	0.00213	0.00087	0.00089	0.00077	0.00077	0.16686	0.1774	0.07643	0.07889	0.0459	0.04549
LSTAR	0.00203	0.00211	0.00087	0.00089	0.00078	0.00077	0.16715	0.17753	0.07643	0.07889	0.04604	0.04549
MLP	0.00201	0.00213	0.00087	0.00089	0.00077	0.00077	0.16501	0.17743	0.07643	0.07889	0.04556	0.0455
JCN	0.00197	0.00208	0.00085	0.00089	0.00064	0.00076	0.16235	0.17794	0.07444	0.07891	0.03808	0.04561
Elman	0.00201	0.00214	0.00012	0.00012	0.00013	0.00013	0.16478	0.17736	0.07694	0.07694	0.08910	0.08910

Transaction cost equals to 0.20%.

[a] Indicates the best result between models in the column of trading rule return when this value is greater than the return on risk-adjusted buy-and-hold strategy.

ratios for all forecast horizons and for one-step-ahead forecasts. Nevertheless, it did not offer significantly better results with respect to the mean Sharpe risk ratio for all models, even though it produces higher mean net returns than does the risk-adjusted buy-and-hold strategy. Indeed, in some cases, a model produces an even higher Sharpe ratio than does the risk-adjusted buy-and-hold strategy (e.g., MLP at 500-forecast horizon, Panel A; and MLP at 200-forecast horizon, Panel B).

5. Conclusions

We compare out-of-sample forecasts of daily returns for the Ibex-35 index, generated by six competing models, namely a linear AR model, the ESTAR and LSTAR smooth transition autoregressive models and three ANN models: MLP, JCN and Elman networks. This comparison was carried out on the basis of various statistical criteria and by assessing the economic value of the predictors. In both cases, we used different forecasting methods (one-step- and multi-step-ahead predictions) and forecast horizons to analyse the robustness of the results.

The results obtained put into question the hypothesis of Efficient Markets, because the random walk model is not a reasonable description of Ibex-35 stock index returns and the ANN technique may be an appropriate way to improve forecasts. However, what are the merits of the ANN model versus its competitors? In the light of the results obtained, the statistical criteria suggest that the out-of-sample ANN forecasts are more accurate, provide a better fit and provide slightly improved predictions than do the AR and STAR models. Furthermore, in some cases, the ANN one-step-ahead forecasts are able to explain the forecast errors of the other models, and the fact that MLP gives the best values. In terms of the economic criteria in the out-of-sample forecasts, we assessed profitability by means of a simple trading strategy known as the filter technique, using a range filter percentage and different trading costs to analyse the robustness of the results against such cost variations. The results indicate a better fit for the ANN models, which offer slightly higher trading profits, in terms of the mean net return and the Sharpe risk-adjusted ratio and also in terms of the one-step-ahead forecasting method (recursive and nonrecursive). Moreover, these results are valid for all the transaction costs analysed. Finally, in the case of multi-step-ahead forecasts, we were unable to determine whether there exist significant trading advantages for the ANNs or for any other model.

After assessing different statistical and economic criteria, we conclude that the return on the Ibex-35 index can be predicted by using ANN models and one-step-ahead forecasts. Therefore, the neural network technique can be said to represent a slight improvement in the investment strategy of short-term agents, with respect to other strategies based on linear and STAR models.

Acknowledgements

We are very grateful to Richard T. Baillie (editor) and to the two anonymous referees for their helpful comments and constructive suggestions. We also thank Simón Sosvilla, Fernando Fernández, Foundation for Applied Economic Research (FEDEA), and research seminar participants at the University of Islas Baleares, Spain (2003). Additionally, the first and third authors acknowledge financial support under Spanish Ministry of Science and Technology grant BEC2001-3777. All errors are the authors' responsibility.

References

Adya, M., Collopy, F., 1998. How effective are neural networks at forecasting and prediction? A review and evaluation. Journal of Forecasting 17, 481–495.

Alexander, S., 1961. Price movements in speculative markets: trends or random walks. Industrial Management Review 2, 7–26.

Boné, R., Crucianu, M., 2002. Multi-step-ahead prediction with neural networks: a review. Publication de l'equipe RFAI, 9′emes rencontres internationals. Approches Connexionnistes en Sciences Économiques et en Gestion, pp. 97–106.

Brock, W., Dechert, W., Scheinkman, J., 1987. A test for independence based on the correlation dimension. University of Wisconsin, Dept. of Economics. Revised in Brock, W., Dechert, W., Scheinkman, J. and LeBaron, B. (1996), Econometric Reviews 15, 197-235.

Chong, Y., Hendry, D., 1986. Econometric evaluation of linear macroeconomic models. Review of Economic Studies 53, 671–690.

Clements, M., Hendry, D., 1993. On the limitations of comparing mean square forecast errors. Journal of Forecasting 12, 617–637.

Clements, M., Smith, J., 1999. A Monte Carlo study of forecasting performance of empirical SETAR models. Journal of Applied Econometrics 14, 123–142.

Corradi, V., Swanson, N., 2001. Bootstrap conditional distribution tests under dynamic misspecification. Mimeo, Rutgers University.

Corradi, V., Swanson, N., 2002. A consistent test for nonlinear out of sample predictive accuracy. Journal of Econometrics 110, 353–381.

De Gooijer, J., Bruin, P., 1998. On forecasting SETAR processes. Statistics and Probability Letters 37, 7–14.

Diebold, F., Mariano, R., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13 (3), 253–263.

Duhoux, M., Suykens, J., de Moor, B., Vandewalle, J., 2001. Improved long-term temperature prediction by chaining of neural networks, International Journal of Neural Systems, vol. 11. World Scientific Publishing, pp. 1–10.

Elman, J.L., 1990. Finding structure in time. Cognitive Science 14, 179–211.

Eitrheim, O., Teräsvirta, T., 1996. Testing the adequacy of smooth transition autoregressive models. Journal of Econometrics 74, 59–75.

Fama, E., Blume, M., 1966. Filter rules and stock market trading profits. Journal of Business 39, 226–241.

Fernández, F., García, D., Sosvilla, S., 1999. Dancing with bulls and bears: nearest-neighbour forecast for the Nikkei index. Japan and the World Economy 11, 395–413.

Fernández, F., Sosvilla, S., Andrada, J., 2003. Technical analysis in foreign exchange markets: evidence from the EMS. Applied Financial Economics 13, 113–122.

García, R., Gençay, R., 2000. Pricing and hedging derivative securities with neural networks and a homogeneity hint. Journal of Econometrics 94, 93–115.

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

Gençay, R., Liu, T., 1997. Nonlinear modeling and prediction with feedforward and recurrent networks. Physica D 108, 119–134.

Granger, C.W.J., 1992. Forecasting stock market prices: lessons for forecasters. International Journal of Forecasting 8, 3–13.

Granger, C.W.J., Teräsvirta, T., 1993. Modelling Nonlinear Economic Relationships. Oxford University Press, Oxford, UK.

Harvey, D., Leybourne, S., Newbold, P., 1998. Tests for forecast encompassing. Journal of Business and Economic Statistics 16, 254–259.

Harvey, D., Leybourne, S., Newbold, P., 1999. Forecast evaluation tests in the presence of ARCH. Journal of Forecasting 18, 435–445.

Hinich, M., Paterson, D., 1985. Evidence of nonlinearity in stock returns. Journal of Business and Economic Statistics 3, 69–77.

Hornik, K., Stinchcombe, M., White, H., 1990. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. Neural Networks 3, 551–560.

Hsieh, D., 1991. Chaos and nonlinear dynamics: applications to financial markets. Journal of Finance 46, 1839–1878.

Klimasauskas, C., 1991. Applying Neural Networks: Part 3. Training a Neural Network, PC-AI, May–June, pp. 20–24.

Kuan, C., White, H., 1994. Artificial neural networks: an econometric perspective. Econometric Reviews 13, 1–91.

Leitch, G., Tanner, J., 1991. Economic forecast evaluation: profits versus the conditional error measures. American Economic Review 81, 580–590.

Lundbergh, S., Teräsvirta, T., 2002. Forecasting with smooth transition autoregressive models. In: Clements, M., Hendry, D. (Eds.), A Companion to Economic Forecasting. Blackwell, pp. 485–509. Chap. 21.

Pemberton, J., 1987. Exact least squares multi-step prediction from nonlinear autoregressive models. Journal of Time Series Analysis 8, 443–448.

Pesaran, M.H., Timmermann, P., 1992. A simple nonparametric test of predictive performance. Journal of Business and Economic Statistics 10, 461–465.

Satchell, S., Timmermann, A., 1995. An assessment of the economic value of non-linear foreign exchange rate forecasts. Journal of Forecasting 14, 477–497.

Scheinkman, J., LeBaron, B., 1989. Nonlinear dynamics and stock returns. Journal of Business 62, 311–338.

Sharpe, W., 1966. Mutual fund performance. Journal of Finance 39, 119–138.

Swanson, N., White, H., 1995. A model-selection approach to assessing the information in the term structure using linear models and artificial neural networks. Journal of Business and Economic Statistics 13 (3), 265–275.

Swanson, N.R., White, H., 1997. Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. International Journal of Forecasting 13, 439–461.

Sweeney, R.J., 1988. Some new filter rule tests: methods and results. Journal of Financial and Quantitative Analysis 23, 285–300.

Teräsvirta, T., 1994. Specification, estimation and evaluation of smooth transition autoregressive models. Journal of the American Statistical Association 89, 208–218.

Teräsvirta, T., Tjøstheim, D., Granger, C.W.J., 1994. Aspects of modelling nonlinear time series. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, vol. IV. Elsevier, pp. 2919–2957. Chap. 48.

West, K., 1996. Asymptotic inference about predictive ability. Econometrica 64, 1067–1084.

White, H., 1988. Economic prediction using neural networks: the case of IBM daily stock returns. Proceedings of the IEEE International Conference on Neural Networks, San Diego, CA.

White, H., 2000. A reality check for data snooping. Econometrica 68, 1097–1126.