Comput Econ (2012) 40:245–264
DOI 10.1007/s10614-011-9288-5

Nonlinearity in Forecasting of High-Frequency Stock Returns

Juan C. Reboredo · José M. Matías · Raquel Garcia-Rubio

Received: 23 September 2010 / Accepted: 1 August 2011 / Published online: 13 August 2011
© Springer Science+Business Media, LLC. 2011

Abstract Using high-frequency S&P 500 data, we examined intraday efficiency by comparing the ability of several nonlinear models to forecast returns for horizons of 5, 10, 30 and 60 min. Taking into account fat tails and volatility dynamics, we compared the forecasting performance of simple random walk and autoregressive models with Markov switching, artificial neural network and support vector machine regression models in terms of both statistical and economic criteria. Our empirical results for out-of-sample forecasts for high and low volatility samples at different time periods provide weak evidence of intraday predictability in terms of statistical criteria, but corroborate the superiority of nonlinear model predictability using economic criteria such as trading rule profitability and value-at-risk calculations.

Keywords Nonlinear models · Intraday returns · Markov switching · Artificial neural networks · Support vector machine regression

JEL Classification C22 · C45 · C52 · C53 · G17

1 Introduction

The prediction of stock market returns has attracted the attention of financial researchers and practitioners for decades. The weak-form of the efficient market hypothesis

J. C. Reboredo (✉)
Department of Economics, University of Santiago de Compostela, Avda. Xoán XXII, s/n, 15782 Santiago de Compostela, Spain
e-mail: juancarlos.reboredo@usc.es

J. M. Matías
Department of Statistics, University of Vigo, Vigo, Spain

R. Garcia-Rubio
Department of Finance, University of Salamanca, Salamanca, Spain

(Fama 1970, 1991) suggests that the price of stocks traded in efficient markets shows martingale behaviour and is thus impossible to predict on the basis of historical price information. Therefore, time series unpredictability or randomness of stock returns provides the basis for the market efficiency test.[1] From a practical point of view, inference regarding market efficiency has notable implications for portfolio and financial managers in designing trading strategies, derivative pricing and financial risk management.

In practice, identifying stock return predictability based on past returns is difficult, particularly for longer time periods. Early tests of market efficiency employed the autocorrelation test and the Lo and MacKinlay variance ratio test (1988) for daily, weekly and monthly stock returns; for example, Liu and He (1991); Lee et al. (2000); Chordia et al. (2005) and Liu (2007) used the autocorrelation test for stocks and currency markets, while Pan et al. (1997); Lee et al. (2000); Chaudhuri and Wu (2003); Patro and Wu (2004) and Bianco and Renò (2006) used the variance ratio test to test the efficiency of stock and exchange rate markets. But, as noted by Hsiec (1991); Mcqueen and Thorley (1991) and Hong and Lee (2003), because these testing procedures assume linearity they only check for serial uncorrelatedness rather than for the martingale property of asset returns.

There is no reason to suppose, at the theoretical level, that stock prices must be intrinsically linear. In fact, nonlinearity in stock returns may arise from price fads, rational speculative bubbles (Mcqueen and Thorley 1991) or the assumption that prices are the result of complex interactions between informed and uninformed traders in the market place. In addition, human error in reasoning or information processing (e.g., information bias or overconfidence) may explain information imperfections in financial markets (Kahneman and Tversky 2000) that may give rise to price nonlinearity. Nonlinearity in stock returns is crucial for unpredictability testing purposes, since a nonlinear time series stock return model could simultaneously have zero autocorrelation and a nonzero mean conditional in its past history, implying predictable nonlinearity-in-mean (Hsiec 1989, 1993). Therefore, both the autocorrelation test and the variance ratio test could wrongly support the martingale hypothesis by failing to detect predictable nonlinearities-in-mean. Nonlinear time series models have been extensively applied in the literature to modelling and forecasting financial time series for daily or monthly frequencies (see, e.g., Meese and Rose 1990, 1991; Gencay 1999; Medeiros et al. 2001; Pérez-Rodríguez et al. 2005; Rapach and Wohar 2006; Bali et al. 2008).

In this article, we focus on intraday market efficiency by examining nonlinear predictability in stock returns observed at different time intervals. Intraday horizons are appropriate for testing market efficiency since we should expect predictability to weaken as trader's actions exploit any profit opportunity emerging from new information, thus forcing prices to adjust to new equilibrium levels. In fact, Lo (2004) has argued that the market does not instantaneously adjust to new and changing economic information; this may lead to temporal serial returns correlation giving rise to a kind of localized predictability, limited in both scope and duration until market participants

[1] From a theoretical point of view, some models showed serial correlation in returns to be consistent with an equilibrium model of asset pricing with risk-averse agents, so serial correlation does not imply a violation of market efficiency (see, e.g., Cecchetti et al. 1990).

learn to take advantage of it. The more efficient the market, the faster new information becomes embodied in prices; predictability should therefore tend to recede over longer horizons. Some tests of market efficiency that examined return serial correlation for daily, weekly or monthly frequencies found weak evidence of serial correlation (see, e.g., Chordia et al. 2005; Timmermann 2008); consequently the efficiency creation process—and thus predictability—must be looked for in intraday trading data. From a practical point of view, intraday forecasting is particularly relevant for practitioners as they can gain high annual returns by exploiting intraday trading strategies. Intraday forecasting horizons are also important for trading desk risk (see Chew 1994). Despite its theoretical and practical relevance, stock return predictability for very short forecasting horizons is under-represented in the academic literature; worthy of mention, however, are studies by Clements and Taylor (2003), Hong et al. (2007) and Wang and Yang (2010).

By examining intraday nonlinear predictability in stock returns, we attempt to contribute to the literature in the following ways. First, we extend the literature by considering the forecasting ability of highly flexible nonlinear models applied to intraday data for different horizons that additionally consider nonlinearity-in-variance. Specifically, we employed a two-state Markov switching (MS) regression model with generalized autoregressive conditional heteroskedasticity (GARCH) errors, an artificial neural network model with GARCH components and a support vector machine (SVM) regression with GARCH errors, adapted from machine learning and whose predictive ability for intraday returns is as yet unexplored. The forecasting ability of these nonlinear models was compared with a simple random walk (RW) model and an autoregressive (AR) model with GARCH errors. Second, we account for fat tails in the probability distributions of the returns—especially pertinent given that high-frequency returns exhibit significant leptokurtosis. Previous studies on intraday returns (e.g., Wang and Yang 2010) fail to model fat tails or time-varying volatility. Third, we provide out-of-sample evidence as it focuses directly on predictability and is particularly important to avoid in-sample overfitting for nonlinear models (see Dacco and Satchell 1999). We also used White's reality check to take data-snooping bias into account, i.e., to check for the superior predictive ability of certain complex models due to chance. Finally, as in Swanson and White (1997), Gencay (1998), Gencay (1999), Hong and Lee (2003) and Wang and Yang (2010), our out-of-sample evidence is discussed in terms of both statistical and economic criteria. From an economic perspective, we evaluated different forecasting models for the profitability of a simple trading rule based on the forecasting abilities of the different models and also the direction of the forecasted price changes, since directional predictability has essential implications for market timing and active intraday asset allocation management. We also compared the forecasting performance of the competing models—as per Brooks and Persand (2003) and Sarma et al. (2003)—by considering a risk management loss function based on calculating value-at-risk (VaR), which, as far as we are aware, has not yet been considered for nonlinearly modelled intraday returns.

We used intraday data for the S&P 500 index for two sample periods with different trend and volatility behaviour, namely, 5 April 2006 to 3 August 2006 and 1 February 2011 to 13 May 2011, for different within-day temporal horizons of 5, 10, 30 and 60 min. Our main empirical findings for out-of-sample one-step-ahead forecasts corroborated the superiority of nonlinear model predictability on the grounds of economic criteria, but provided somewhat limited evidence for predictability with respect to simple linear models based on traditional statistical criteria. For statistical criteria such as mean squared forecast error or out-of-sample log-likelihood evaluations, weak evidence of predictability was found for the high volatility sample period in terms of log-likelihood; however, according to White (2000) reality check results, no evidence was found of clear-cut superiority of the nonlinear models over the simple RW model in terms of mean squared forecast error. More robust evidence of predictability was found on economic grounds. Trading returns arising from a simple trading strategy based on an SVM forecast for a 60-min period yielded better out-of-sample profitability for both sample periods. We also found clear evidence of the superiority of nonlinear model over linear models on the basis on the risk management loss function. However, we found no significant differences in market timing ability between linear and nonlinear models.

The rest of the paper is laid out as follows: Section 2 describes the different nonlinear forecasting time series models; Sect. 3 describes our data and examines the predictive capacity of several models for different intraday time periods; and finally, Sect. 4 contains our conclusions.

2 Forecasting Models

Let r_t be a continuous intraday stock return with conditional mean $E(r_t|I_{t-1})$, so that

$$r_t = E\left(r_t|I_{t-1}\right) + \varepsilon_t, \tag{1}$$

where $I_{t-1} = \{r_{t-1}, r_{t-2}, \ldots\}$ is the information set at time $t-1$ on which forecasts are based, i.e., the σ-algebra induced by all observed variables at time $t-1$. ε_t is the forecast error which is assumed to follow a skewed Student-t distribution to cope with fat tails in the return distribution:

$$\sqrt{\frac{\upsilon}{\sigma_t^2(\upsilon-2)}}\,\varepsilon_t \sim \text{iid} \quad t_\upsilon, \tag{2}$$

where υ are the degrees of freedom and σ_t^2 is the conditional variance of ε_t, $E\left(\varepsilon_t^2\,|I_{t-1}\right) = \sigma_t^2$; this evolves, accordIng to a GARCH (1,1) process, into:

$$\sigma_t^2 = \omega + \beta\,\sigma_{t-1}^2 + \alpha\,\varepsilon_{t-1}^2, \tag{3}$$

where ω is a constant; σ_{t-1}^2 is the previous period's forecast error variance (the GARCH component); ε_{t-1} is news on volatility from previous periods (the autoregressive conditional heteroskedasticity (ARCH) component). To ensure positivity of the conditional variance, the parameters of the conditional volatility model, ω, β and α, are assumed to be positive. This conditional volatility model is able to capture intraday time-varying volatility, volatility clustering and, to produce a forecast variance, $\hat{\sigma}_t^2$, along with a forecast error such that the standardized residuals, $\hat{\varepsilon}_t/\hat{\sigma}_t$ are homoskedastic and independent.

We applied a number of flexible nonlinear models that can capture potential nonlinearities in the conditional mean of stock returns and assume that nonlinearities-in-variance are well represented by (3). As a benchmark we took the RW model, $r_t = \mu + \varepsilon_t$, in which the forecast error is assumed to be homoskedastic. The set of models for the conditional mean in (1) included a simple linear autorregressive model, AR(p) (p was selected to minimize the Schwarz criterion), the two-state MS regression, a feedforward artificial neural network (ANN), an SVM regression and a combination of these models (denoted COMB). Below we describe briefly the nonlinear models used in this study.

2.1 Markov Switching GARCH Model

The MS model introduced by Hamilton (1989) allows for some or all the parameters of the conditional mean model to switch across different regimes—e.g., in periods of high or low return volatility—according to a Markov process governed by a state variable denoted by S_t. For simplicity sake, we considered a two-state regime and adopted a similar approach to Gray (1996) and Dueker (1997). Thus, the conditional mean can be written as:

$$E\left(r_t|I_{t-1}\right) = \left(\gamma_1 + \delta_1 r_{t-1}\right) S_t + \left(\gamma_0 + \delta_0 r_{t-1}\right)\left(1 - S_t\right), \tag{4}$$

where $S_t \in \{0, 1\}$ $\forall t$ is the latent Markov chain of order one with transition probabilities $\Pr\left(S_t = 0|S_{t-1} = 0\right) = p$ and $\Pr\left(S_t = 1|S_{t-1} = 1\right) = q$. Therefore, nonlinearities-in-mean are considered to be a mix of distributions with different dynamic properties, where the value of the variable is drawn from the most likely unobservable state. The ergodic or unconditional probability of being in state one is given by $(1-q)/(2-p-q)$.

Additionally, the innovation term in (1), ε_t, has a conditional variance that switches from one GARCH process to another:

$$\sigma_t^2 = \left(\omega_1 + \beta_1\,\sigma_{t-1}^2 + \alpha_1\,\varepsilon_{t-1}^2\right) S_t + \left(\omega_0 + \beta_0\,\sigma_{t-1}^2 + \alpha_0\,\varepsilon_{t-1}^2\right)\left(1 - S_t\right), \tag{5}$$

where σ_{t-1}^2 is a state-independent average of past conditional variances; this is because, in a regime-switching framework, a GARCH model with state-dependent past conditional variances is unfeasible as the number of (unobserved) regime paths would grow exponentially in line with sample size. To circumvent the problem of path dependence, we adopted the suggestion of Gray (1996)—which was also followed by Dueker (1997)—consisting of the use of the conditional expectation of the past variance to integrate out the unobserved regimes. Thus:

$$\begin{aligned}\sigma_{t-1}^2 = E_{t-2}\left(\sigma_{t-1}^2\right) &= p_{0,t-1}\left[\left(\mu_{t-1}^{(0)}\right)^2 + \sigma_{t-1}^{2(0)}\right] + \left(1 - p_{0,t-1}\right)\left[\left(\mu_{t-1}^{(1)}\right)^2 + \sigma_{t-1}^{2(1)}\right] \\ &\quad - \left[p_{0,t-1}\mu_{t-1}^{(0)} + \left(1 - p_{0,t-1}\right)\mu_{t-1}^{(1)}\right]^2,\end{aligned} \tag{6}$$

where $p_{0,t-1} = \Pr(S_{t-1} = 0|I_{t-2})$, $\mu_{t-1}^{(j)} = \gamma_j + \delta_j r_{t-2}$, $\sigma_{t-1}^{2(j)} = \left(\omega_j + \beta_j \sigma_{t-2}^2 + \alpha_j \varepsilon_{t-2}^2\right)$ for $j = 0,1$. Moreover, $\varepsilon_{t-1} = r_{t-1} - E_{t-2}(r_{t-1}|I_{t-2})$, where $E_{t-2}(r_{t-1}|I_{t-2}) = p_{0,t-1}\mu_{t-1}^{(0)} + \left(1 - p_{0,t-1}\right)\mu_{t-1}^{(1)}$. Equation 5 allows both the conditional variance to switch and the unconditional variance of the GARCH process, $\omega/1 - \alpha - \beta$, to have two different regimes. Equation 6 enables one-step-ahead volatility forecasting to be made, whereas the one-step-ahead forecast for the conditional mean can be calculated as $E_{t-1}(r_t|I_{t-1}) = p_{0,t}\mu_t^{(0)} + \left(1 - p_{0,t}\right)\mu_t^{(1)}$.

To estimate the model parameters, adopted in practice in the Markov regime-switching literature is the maximum likelihood estimation. The log-likelihood function can be written as:

$$\ell = \sum_{t=1}^{T} \log\left[p_{0,t} f(r_t|S_t = 0) + \left(1 - p_{0,t}\right) f(r_t|S_t = 1)\right], \tag{7}$$

where $f(r_t|S_t = j)$ is the conditional distribution, given that regime j occurs at time t and $p_{0,t}$ is the probability of being in regime 0 at time t, given by:

$$p_{0,t} = \Pr(S_t = 0|I_{t-1}) = \sum_{j=0}^{1} \Pr(S_t = 0|S_{t-1} = j)\left[\frac{f(r_{t-1}|S_{t-1} = j)\, p_{i,t-1}}{\sum_{k=0}^{1} f(r_{t-1}|S_{t-1} = k)\, p_{k,t-1}}\right]. \tag{8}$$

Further details on the maximization of (7) can be found in Hamilton (1994).

2.2 Artificial Neural Network Model

Neural networks are very flexible nonlinear models which, due to their approximation capacity, can be used under a nonparametric philosophy, with the model adapting to the underlying data structure (see, e.g., Haykin 1999; Hastie et al. 2001). This approximation capacity is due to a multilayer structure composed of at least one intermediate hidden layer between the input and output that combines many basic nonlinear functions. This hidden layer makes a previous nonlinear transformation of the data so as to facilitate resolution of the problem in hand (regression, classification, etc.). Given the aims described in the introduction, the neural networks used for this research are regression models.

The neural networks most frequently used for regression are the multilayer perceptron (MLP) network and the radial basis function (RBF) network with a hidden layer. Both models have been used to capture a wide variety of nonlinear patterns (see Hornik et al. 1990), for which they generate good predictions (see, e.g., Swanson and White 1995, 1997; Gencay 1999) thanks to their universal approximation property (see, e.g., Leshno et al. 1993), in other words, their capacity to approximate any continuous function provided they have a sufficient number of hidden units.

The basic difference between the two kinds of networks is the type of basis function used in the hidden layer: radial for the RBF, based on the input space norm or distance, and projection for the MLP, based on the inner product of the input space (see Eq. 9 below).

The SVMs (see the next section) are also a linear combination of radial-type basis functions. They have been demonstrated to be superior to the RBFs, due primarily to their more efficient training algorithm and greater parsimony. For this reason, we preferred to use SVMs to represent the radial basis models and MLP networks to represent the projection philosophy.

The MLP model has been extensively used for economic and financial forecasting purposes. The MLP(p, q) has the following specification:

$$f(\mathrm{r}_t; \theta) = \sum_{j=1}^{q} c_j \psi(w_{j0} + \mathrm{w}_j' \mathrm{r}_t) + c_0, \tag{9}$$

where θ denotes all the model parameters, p is the dimension of the input space, q is the number of hidden layers, ψ is a sigmoid function (typically, a logistic or hyperbolic tangent function), $w_{j0} \in \mathbb{R}$, $\mathrm{w}_j \in \mathbb{R}^p$ are the parameters of the jth hidden unit, $\mathrm{w}_j' \mathrm{r}_\mathrm{t}$ is the inner product between w_j and r_t and c_j, $j = 0, 1, \ldots, q$ are the output coefficients for the hidden units. Following Gencay (1999) and Campbell et al. (1997) we chose the logistic function for ψ, given that it has a threshold behaviour that is useful in capturing many kinds of nonlinearity patterns, especially for extreme values. In order to ensure consistency in a stochastic environment, in addition to considering the error associated with the approximation of the regression function via a finite number of parameters, we also needed to consider the estimation error arising from the use of a limited quantity of data. The consistency of the MLP neural networks for different hypotheses was thus obtained (see Krzyzak et al. 1996; Fine 1999), specifically for dependent observations (Chen and White 1999).

When the series responds to an AR model, we have:

$$r_\mathrm{t} = f(\mathbf{r}_{t-1}; \theta) + \varepsilon_\mathrm{t}, \tag{10}$$

where $f(\mathbf{r}_{t-1}; \theta)$ is an MLP neural network and ε_t is the error term described in Eqs. 1–3. Trapletti et al. (2000) demonstrated the stationarity and strongly mixing nature of the series $\{\mathrm{r}_\mathrm{t}\}$, as also the consistency and asymptotic normality of the least squares estimator for a hypothesis that ensures the identifiability of the network parameters (Hwang and Ding 1997).

Our use of the MLP networks in this research is novel in two ways: (1) they are combined with a GARCH model (Eq. 3) for the variance; (2) the parameters are estimated jointly using the conjugate gradient algorithm (see, e.g., Press et al. 1992) and maximum likelihood as the estimation criterion (rather than the classical minimum squared error) for the MLP component of the mean.

2.3 Support Vector Machine for Regression

SVMs for regression (Vapnik 1998; Schölkopff and Smola 2002) are linear models obtained in a new feature space $\mathfrak{X}$ as a result of a transformation $\varphi : \mathbb{R}^p \rightarrow \mathfrak{X} \subset \mathbb{R}^q$

of the input space. Therefore, SVMs for regression have the following general formulation[2]:

$$f(\mathbf{r}_t, \theta) = \langle \mathbf{w}, \varphi(\mathbf{r}_t) \rangle + b,$$

where θ denotes all the model parameters, $\mathbf{w} \in \mathbb{R}^q$ are the linear model coefficients and b is the independent term and where $\varphi(\mathbf{r}_t)$ is the image of $\mathbf{r}_\mathrm{t}$ in the new transformed space $\mathfrak{X}$.

As with the MLP model, when the series responds to an AR model, we have the SVM(p) model:

$$r_\mathrm{t} = f(\mathbf{r}_{t-1}; \theta) + \varepsilon_t, \tag{11}$$

where $f(\mathbf{r}_{t-1}; \theta)$ is an SVM, $\mathbf{r}_{t-1} \in \mathbb{R}^p$ and ε_t is the error term described in (1)–(3).

Given a sample, the parameters $\mathbf{w}$, b in the SVM are estimated as the solution to the following regularization problem:

$$\min_{\mathbf{w},b} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{t=p+1}^{T} \ell(r_\mathrm{t}, f(\mathbf{r}_{t-1})) \right\}, \tag{12}$$

where C is a regularizing constant and ℓ is the δ-insensitive loss:

$$\ell(r_t, f(\mathbf{r}_{t-1})) = |r_t - f(\mathbf{r}_{t-1})|_\delta = \max\{0, (|r_t - f(\mathbf{r}_{t-1})| - \delta)\}$$

where δ is a constant that prevents overfitting of the data, with any error less than δ not counting as such in the objective function. C, meanwhile, is a parameter that regularizes solution complexity, determined by the size of the components of $\mathbf{w}$ through its norm $\|\mathbf{w}\|$. The larger C, the more important the data fit, and the smaller C, the more important it is to limit complexity, even if a poorer fit results.

Equation 12 is resolved by a dual formulation, resulting in a squared problem with restrictions and with a unique solution. The only solution to the problem in Eq. 12 is a linear combination of a set S of key points of the sample (the support vectors), $\mathbf{w} = \sum_{s\in S} \alpha_s \varphi(\mathbf{r}_s)$, in such a way that the SVM results as:

$$f(\mathbf{r}_{t-1}; \theta) = \sum_{s\in S} \alpha_\mathrm{t} \langle \varphi(\mathbf{r}_{t-1}), \varphi(\mathbf{r}_s) \rangle + b. \tag{13}$$

For the above expression to be useful, we need to determine the transformation φ in the input space. This transformation requires that the problem be transferred to a larger dimension space than the input space $\mathfrak{X}$ (possibly infinite), to ensure that the linear model in this space is adequate to resolving the problem. The selection of φ and the computational burden associated with calculating the inner product implied by Eq. 13

[2] The notation of the inner product using angled brackets $\langle \mathbf{w}, \varphi(\mathbf{r}_t) \rangle$is traditional in these models and so is retained. In reality,$\langle \mathbf{w}, \varphi(\mathbf{r}_\mathrm{t}) \rangle = \mathbf{w}' \varphi(\mathbf{r}_t)$, following the notation used to date.

are both simplified by defining the inner product in terms of a positive definite function (kernel):

$$\langle \varphi(\mathbf{r}_{t-1}), \varphi(\mathbf{r}_s) \rangle = k(\mathbf{r}_{t-1}, \mathbf{r}_s). \tag{14}$$

Equation 13 thus results in:

$$f(\mathbf{r}_{t-1};\theta) = \sum_{s\in S} \alpha_{\mathrm{t}} \langle \varphi(\mathbf{r}_{t-1}), \varphi(\mathbf{r}_s) \rangle + b = \sum_{s\in S} \alpha_s k(\mathbf{r}_{t-1}, \mathbf{r}_s) + b.$$

When a priori information regarding the selection of this kernel is lacking, it is normal to use the Gaussian kernel, given its good general behaviour:

$$k(\mathbf{r}_{t-1}, \mathbf{r}_s) = \exp\left\{ -\frac{1}{2\sigma^2} \left\| \mathbf{r}_{t-1} - \mathbf{r}_s \right\|^2 \right\} \tag{15}$$

where σ acts as a window parameter and determines the complexity of the model (although any positive definite function is possible). With a kernel of this nature, SVMs share the general form of RBF neural networks, which are universal approximators to continuous or integrable functions (see, e.g., Park and Sandberg 1993). Consistency results also exist (see Bousquet and Elisseeff 2002; Steinwart 2002).

Equation 12 has a unique solution for each value of C, δ and σ (Gaussian kernel parameter). Hence, these parameters have to be selected using some model selection procedure. Cross-validation is typically used and was the method used for this research. We combined an SVM model for the mean with a GARCH model for the variance estimated with the errors produced by the former. This sequential estimation is inevitable if the SVM philosophy—obtaining a unique solution and identifying the support vectors—is to be preserved.

2.4 Forecasting Combination

Combining forecasts from different models over a single forecasting model can sometimes improve forecasting accuracy (see, e.g., Clemen 1989; Yang 2004). The difficult in achieving a combined forecast is to determine the weighting combination of forecasts. So, given a combined model:

$$\mathrm{r_t} = \sum_{\mathrm{h}=1}^{\mathrm{H}} \omega_{\mathrm{h}} \hat{\mathrm{r}}_{\mathrm{t,h}}, \tag{16}$$

where $\hat{\mathrm{r}}_{\mathrm{t,h}}$ is the forecast of model h for time t, there exist several ways to determine the value of the weight parameter ω_{h} for each model h. Here, we adopt a simple approach that consists of performing a regression of the observed values of $\mathrm{r_t}$ on in-sample forecasts from the h models for the purpose of obtaining the regression coefficients, i.e., the weights for each model that will be used as weights in the out-of-sample forecasting combination (see Granger and Ramanathan 1984).

3 Empirical Results

3.1 Data

We used intraday data for the S&P 500 index, obtained from Bloomberg, for two different periods: 5 April 2006 to 3 August 2006 and 1 February 2011 to 13 May 2011. We constructed 5-, 10-, 30- and 60-min intraday price time intervals; for each interval the value of the index was taken as the index value closest to the end of the time interval (e.g., for the 5-min interval, the closest value to 9:35 am, 9:40 am, etc). Intraday continuous stock returns for the different time intervals were then computed as the first difference for the logarithmic stock price index.

Table 1 provides descriptive statistics for intraday returns for different time intervals and periods. Average returns were different for the two periods and time intervals, decreasing as the time interval enlarged for the first period and increasing for the second period. Volatility was clearly higher in the second compared to the first period. Likewise, the maximum and minimum values for both periods indicate that price fluctuations were greater in the second compared to the first period. Since the means of the return series were very small relative to the standard deviations, there was no significant trend in the data. The negative values of the skewness statistic for the second period contrast with positive values for the first period, suggesting different probabilities for large decreases in returns for both periods. The high values for the kurtosis statistic in both periods, notably in the second, suggest that the returns distribution has a fat tail, so the fat-tail distribution in Eq. 2 is necessary to characterize the return conditional distribution. In fact, the Jarque-Bera test strongly rejected the normality of unconditional distribution for all the series. Moreover, the Ljung-Box statistic

Table 1 Descriptive statistics for intraday returns

	5 April 2006 to 3 August 2006				1 February 2011 to 13 May 2011			
	5 mins	10 mins	30 mins	60 mins	5 mins	10 mins	30 mins	60 mins
Mean ($\times 10^3$)	−0.003	−0.007	−0.021	−0.039	0.006	0.011	0.032	0.066
Std. Dev. ($\times 10^3$)	0.804	1.153	1.981	2.855	0.960	1.341	2.248	3.328
Max ($\times 10^3$)	5.964	7.040	12.376	14.122	16.484	16.852	15.560	15.478
Min ($\times 10^3$)	−5.852	−6.717	−7.850	−8.847	−19.654	−18.537	−18.050	−17.336
Skewness	0.447	0.400	0.249	0.425	−0.651	−0.278	−0.308	−0.310
Kurtosis	9.754	7.160	6.043	5.468	102.085	49.239	18.652	10.562
Jarque−Bera	12833*	2449*	432*	142*	22973*	9500*	9559*	1033*
$Q(20)$	22.96*	36.43	29.56*	23.50*	22.96*	36.43	29.56*	23.50*
ARCH−LM	0.03	0.10	2.36*	7.60*	0.03	0.10	2.36*	7.60*
Observations (n)	6635	3275	1091	503	5615	2807	935	431

Note. Data are intraday S&P 500 data for the periods indicated in the columns heads. Jarque-Bera is the χ^2 statistic for the test of normality. $Q(k)$ is the Ljung-Box statistics for serial correlation in the squared returns computed with k lags. ARCH-LM is Engel's Lagrange multiplier test for heteroskedasticity, conducted using 20 lags. An asterisk (*) indicates a rejection of the null hypothesis at the 5% level indicated no evidence of serial correlation except for the 10-min returns. Finally, the ARCH-Lagrange multiplier (ARCH-LM) statistic indicated that ARCH effects were likely in return series at long time intervals, but not for short time interval returns. Obviously, the size of the sample decreased as the time interval grew—from more than 5500 items for a 5-min interval to less than 500 for a 60-min interval.

Hence, in our data set we have two distinct periods for the analysis of the predictability of stock returns using nonlinear models: a relative tranquil market period with low volatility and a price drop (−2.12%) and a waved market period with high volatility and a price increase (3.25%). These two sets of data allow us to explore whether nonlinear modelling predictability evidence is stronger for volatile market conditions or for a bear market.

3.2 Models, Estimates and Evaluation

Following on from the previous sections, the following models were analysed:

- RW model
- AR(1)-GARCH(1,1) MS model with two states (following Grey, 1996), estimated using maximum likelihood.
- Heteroskedastic MLP(1,2)-GARCH(1,1) model, estimated jointly using maximum likelihood and the conjugate gradient method. Therefore, the MLPs possess a priori $q = 2$ hidden units. This decision was based on an exploratory in-sample analysis goodness-of-fit for different values of this parameter and on an analysis of the scatter plots for the data (r_t vs. r_{t-1}) that suggest the degree of nonlinearity in the data. The model was estimated jointly using maximum likelihood and the conjugate gradient method. The initial values of the parameters for the joint model were obtained by separately estimating the two components of the model: the MLP(1,2) component for the trend was estimated using the standard back-propagation algorithm and the GARCH(1,1) component by the errors obtained for the former. We subsequently applied the conjugate gradient method to all the parameters of the joint heteroskedastic model (MLP+GARCH), setting a lower threshold for changes in the solution as the stop criterion.
- Heteroskedastic SVM(1)-GARCH(1,1) model, with the two components estimated separately and sequentially: first the SVM(1) model was estimated for the mean and then the GARCH(1,1) component was estimated with the residuals. As indicated, there are no known algorithms for joint estimation for the heteroskedastic SVM models, nor is their formulation obvious unless the SVM philosophy based on the following criteria is contravened: parsimonious structure based on the support vector concept, δ-insensitive loss function to control the fit to the data and quadratic optimization problem with a unique solution. Ten-fold cross-validation was used to select the SVM model and for the joint selection of its C and σ parameters, with the parameter δ set as the standard 0.1 value after normalizing the data (the goodness of that decision was first tested with the estimation sample).
- Model combining the above models using linear regression.

Therefore, all the trend models were autoregressive with one lag ($p = 1$, except, logically, the RW model) and all the variance models had a GARCH(1,1) structure.

Regarding the likelihood, it was observed that the best results obtained for the estimation sample were for the Student-t models with values between 3 and 5 degrees of freedom according to the series frequency and the period. In the interest of simplicity in analysing the results, it was decided to set this parameter as 5 degrees of freedom.

The forecasting ability of the different models presented above was assessed on the basis of out-of-sample one-step-ahead forecasting. We took the first 80% of the data as a training data set in order to estimate the models for each of the two samples and the corresponding time intervals. Once the models were estimated, out-of-sample forecasts for the last 20% of the data for the conditional mean and the conditional variance were generated without re-estimating the models obtained in the training set.

3.3 Results

Forecast assessment of competing models is a fundamental step in any forecasting exercise. However, evaluation implies the use of a loss function which can be postulated on statistical grounds or in terms of the loss faced by the final user of the forecasts. As there is no universally acceptable criterion for evaluating the accuracy of a forecast, this study evaluates forecast accuracy using both statistical and economic criteria as described below.

3.3.1 Statistical Criteria

For the sequence of out-of-sample-forecasts, we employed two statistical loss functions. First, we test for the out-of-sample best model using the out-of-sample log-likelihood, which mitigates in-sample overfitting of nonlinear models in selecting satisfactory models. Second, we employed the commonly used the mean squared forecast error metric to check for accuracy and reliable forecasts for the sequence of out-of-sample forecasts arising from the different classes of model. For these two criteria, we compared the out-of-sample performance of competing models for different intraday frequencies with the benchmark RW model—representing linear unpredictability—using White's reality check for the significance of the best statistic obtained. White's reality check is a nonparametric hypothesis test designed to test the null hypothesis that the best model among alternative specifications has no predictive superiority over the benchmark according to a pre-specified loss function. This test accounts for data-snooping bias and thus avoids misleading inferences generated by chance for genuinely good results. As the distribution of the White test statistic is not unique under the null, consistent estimates of the *p*-values were obtained via bootstrapping.

Table 2 reports the value of the out-of-sample log-likelihood for each model for the different time intervals and time periods and the *p*-values on comparing each model with the benchmark. The log-likelihood results were dissimilar depending on the time period considered. For the low volatility period, except for the MS model for the 5- and 10-min intervals, all the models performed equally well given that the reality check was unable to reject the null hypothesis. This means that neither the AR model nor any competing nonlinear and combined models are significantly better than the benchmark on the basis of out-of-sample likelihood. For the high volatility period,

Table 2 Out-of-sample log-likelihood comparison for all models and periods

	5 April 2006 to 3 August 2006				1 February 2011 to 13 May 2011			
	5 mins	10 mins	30 mins	60 mins	5 mins	10 mins	30 mins	60 mins
RW	7654.9	3522.2	1046.2	444.2	6512.2	3051.5	918.1	378.0
AR	7649.5	3526.8	1041.5	442.8	6592.8	3084.6	902.8	373.9
	(0.69)	(0.29)	(0.92)	(0.73)	(0.01)	(0.01)	(0.99)	(0.94)
MS	7681.2	3529.8	1045.8	443.7	6561.2	3060.3	920.9	378.4
	(0.01)	(0.11)	(0.55)	(0.60)	(0.02)	(0.02)	(0.03)	(0.15)
ANN	7651.4	3525.4	1039.9	442.2	6593.1	3085.7	907.5	353.0
	(0.63)	(0.34)	(0.94)	(0.77)	(0.02)	(0.06)	(0.99)	(0.99)
SVM	7652.0	3526.2	1041.5	443.3	6511.9	3076.6	912.7	373.6
	(0.61)	(0.31)	(0.92)	(0.61)	(0.50)	(0.04)	(0.99)	(0.94)
COMB	7659.3	3518.4	1039.5	442.4	6567.8	3075.9	923.1	378.5
	(0.17)	(0.88)	(0.93)	(0.75)	(0.02)	(0.07)	(0.08)	(0.40)

Note. RW, AR, MS, ANN, SVM and COMB denote random walk model, autoregressive model, Markov switching regression model, artificial neural network, support vector machine regression and combined forecast model, respectively. In parenthesis are the bootstrap p-values for comparing a single model with the benchmark—the RW model—using the White reality check with 2000 bootstrap replications

nonlinear models outperformed the benchmark for short time horizons of 5 min, with the exception of the SVM model, and of 10 min at the 5 and 10% significance levels. For longer intraday time horizons, only the MS model remained significant at the 5% level for the 30-min interval, whereas the benchmark was not outperformed by any competing model for the 60-min interval.

Table 3 shows the results for the mean squared forecast error for all models for different time intervals and periods, along with the p-values for White's reality check. As for the log-likelihood, the results for the low volatility period were quite similar across models: although the mean squared error was, in general, lower for the nonlinear models, according to the p-values for the reality check, we were unable to reject the null hypothesis of no predictive superiority of the more sophisticated models over the simple RW benchmark. Similarly, for the high volatility period, the mean squared errors were lower for nonlinear models than for the benchmark, mainly for the shorter time intervals of 5 and 10 min. However, according to the p-values for the reality check, we were again unable to reject the null of no predictive superiority of the sophisticated models over the simple RW benchmark. The lack of success of the nonlinear models in predicting intraday data in terms of mean squared error is consistent with previous empirical results (see, e.g., Wang and Yang 2010) and provides evidence that the market is efficient over short time horizons.

3.3.2 Economic Criteria

As statistical and economic forecast evaluations may lead to different outcomes (see, e.g., Leitch and Tanner 1991; Satchell and Timmerman 1995; Wang and Yang 2010),

Table 3 Out-of-sample mean squared forecast error comparison for all models and periods

	5 April 2006 to 3 August 2006				1 February 2011 to 13 May 2011			
	5 mins	10 mins	30 mins	60 mins	5 mins	10 mins	30 mins	60 mins
RW	0.6811	1.4269	**4.5890**	10.8138	0.7107	1.3882	3.4177	10.8945
AR	0.6783	1.4295	4.5973	**10.7423**	0.6942	1.3469	3.4862	10.9090
	(0.95)	(0.78)	(0.72)	(0.33)	(0.95)	(0.78)	(0.72)	(0.33)
MS	0.6755	1.4272	4.5893	10.7612	**0.6937**	1.3495	**3.4124**	10.8868
	(0.54)	(0.54)	(0.54)	(0.41)	(0.54)	(0.54)	(0.54)	(0.41)
ANN	0.6750	1.4324	4.7258	10.8584	0.6938	**1.3467**	3.5067	10.9091
	(0.27)	(0.85)	(0.81)	(0.58)	(0.27)	(0.85)	(0.81)	(0.58)
SVM	**0.6748**	**1.4270**	4.5911	10.7458	0.7612	1.4133	3.4740	**10.7381**
	(0.23)	(0.51)	(0.65)	(0.44)	(0.23)	(0.51)	(0.65)	(0.44)
COMB	0.6751	1.4331	4.6354	10.860	0.6942	1.3551	3.5983	11.080
	(0.28)	(0.88)	(0.73)	(0.58)	(0.28)	(0.88)	(0.73)	(0.58)

Note. RW, AR, MS, ANN, SVM and COMB denote random walk model, autoregressive model, Markov switching regression model, artificial neural network, support vector machine regression and combined forecast model, respectively. In bold is the minimum value for the mean squared error for each time interval. In parenthesis are the bootstrap p-values for comparing a single model with the benchmark—the RW model—using the White reality check with 2000 bootstrap replications

we assessed forecast performance using three economic criteria of particular relevance to profit-maximizing investors and financial risk managers. First, we considered the trading return for a simple investment rule driven by the return forecast for the out-of-sample one-ahead-period of the competing models. For an initial endowment of 100 monetary units, an investor had to decide whether to maintain this wealth in cash or in assets, depending on whether the return predicted for the next period exceeded a threshold given by the transaction costs. Thus, when the share price was expected to fall below this threshold, shares were sold if the agent held assets; and when the share price was expected to rise above this threshold, shares were purchased if the agent held cash or were kept if the agent held assets. Otherwise, there was no portfolio movement. The threshold was determined by the transaction cost, which was assumed to be low for intraday transactions, as otherwise commissions would erode profits. The value of this cost, which depends on the specific form of the tariff to be paid for transactions, was assumed to be 1 basis point of the value of the transaction. Note that this cost determined the size of movement of the forecast price in order to trigger a trading action and, consequently, the number of transactions in the simulation exercise. Table 4 reports the final portfolio value that the investor would obtain by considering different intraday time periods in the two subsamples. For the low volatility period, nonlinear models outperformed the benchmark RW or the linear AR model; thus, for the 5- 10- and 30-min intervals, the SVM, MS and ANN models, respectively, gave better returns than any other model, even though profitability was not substantially higher than for the simple linear models. However, for a 1-hour forecasting period, the SVM model, with a return of 1.58%, performed slightly better than the ANN and COMB models in terms of profitability; on an annual basis, this became 23.4%, since 20% of the data

Table 4 Cumulative returns for different forecasting models

	5 April 2006 to 3 August 2006				1 February 2011 to 13 May 2011			
	5 mins	10 mins	30 mins	60 mins	5 mins	10 mins	30 mins	60 mins
RW	100	100	100	100	100	100	100	100
AR	99.54	100.06	100.25	101.38	100.02	98.74	96.99	100.09
MS	100.00	**100.19**	100.00	101.07	100.00	100.00	100.00	101.00
ANN	99.65	99.50	**100.37**	101.53	100.00	**100.17**	97.15	99.87
SVM	**100.43**	100.11	100.12	**101.58**	**100.55**	98.36	**100.82**	**101.46**
COMB	100.12	99.30	99.90	101.51	99.90	100.00	100.01	99.70

Note. RW, AR, MS, ANN, SVM and COMB denote random walk model, autoregressive model, Markov switching regression model, artificial neural network, support vector machine regression and combined forecast model, respectively. In bold is the maximum value achievable from the best performing model for each time interval on the basis of forecasting profitability

in the out-of-sample period meant that 17 forecasting days were considered. For the high volatility period, the SVM model was predominant in terms of profitability for all the intraday periods except for the 10-min interval, for which the ANN performed better. Profitability was lower for shorter time periods than for longer periods; even so it was slightly higher than the one obtained in the low volatility period. For the 1-h period, the return for the SVM model was 1.46 or 24.5% on annual basis, since 20% of the data in the out-of-sample period meant that 15 forecasting days were considered. Overall, these results show that although return profitability was low for the linear models, consistent with RW price behaviour, nonlinear modelling of expected returns generates considerable profits, even for periods of up to 60 min. These conclusions remained similar for both the low and high volatility periods studied.

Second, as an alternative economic criterion we considered the direction of the forecasted price changes; for investors, the directional predictability ability of a model has practical implications for market timing and asset allocation management. Table 5 reports the results for the proportion of times that the signs of the returns were correctly forecasted, along with the results of the directional accuracy test proposed by Pesaran and Timmermann (1992). The results were conclusive: no single model was better at forecasting the direction of price changes for the two sample periods and the different intraday horizons. The directional accuracy test was unable to reject the null hypothesis of correct directional prediction for any model or time period.

Third, we evaluated the competing models by quantifying VaR as a measure of the market risk of a portfolio, giving likely losses from future market fluctuations. For a model h, the VaR for the intraday horizon t for a one-step-ahead forecast at the α % significance level is given by:

$$\mathrm{VaR}_{\mathrm{t}}^{\mathrm{h}}(\alpha) = \hat{\mathrm{E}}\left(r_{\mathrm{t+1,h}}\right) + t_{\upsilon}^{-1}(\alpha)\,\hat{\sigma}_{\mathrm{t+1,h}}, \tag{17}$$

where $\hat{\mathrm{E}}\left(r_{\mathrm{t+1,h}}\right)$ and $\hat{\sigma}_{\mathrm{t+1,h}}$ are the mean and standard deviation return forecast by the model h and where t_{υ} is the cumulative t distribution function with υ degrees of

Table 5 Out-of-sample directional predictability for all models and periods

	5 April 2006 to 3 August 2006				1 February 2011 to 13 May 2011			
	5 mins	10 mins	30 mins	60 mins	5 mins	10 mins	30 mins	60 mins
RW	0.50	0.50	0.51	0.50	0.50	0.51	0.50	0.51
	(0.08)	(−0.25)	(0.47)	(−0.19)	(0.24)	(0.83)	(0.24)	(0.53)
AR	0.49	0.51	0.50	0.48	0.52	0.52	0.47	0.52
	(0.50)	(0.65)	(−0.12)	(−0.88)	(0.97)	(0.81)	(−1.38)	(0.97)
MS	0.49	0.55	0.49	0.50	0.51	0.47	0.51	0.47
	(−0.39)	(1.32)	(−0.26)	(0.08)	(0.29)	(−0.86)	(0.29)	(−0.86)
ANN	0.48	0.50	0.50	0.50	0.48	0.48	0.48	0.48
	(−0.48)	(−0.18)	(−0.10)	(−0.16)	(−0.39)	(−0.33)	(−0.34)	(−0.36)
SVM	0.50	0.50	0.51	0.50	0.50	0.51	0.50	0.51
	(0.08)	(−0.25)	(0.47)	(−0.19)	(0.24)	(0.83)	(0.24)	(0.53)
COMB	0.49	0.51	0.50	0.48	0.52	0.52	0.47	0.52
	(−0.50)	(0.65)	(−0.12)	(−0.88)	(0.97)	(0.81)	(−1.38)	(0.97)

Note. RW, AR, MS, ANN, SVM and COMB denote random walk model, autoregressive model, Markov switching regression model, artificial neural network, support vector machine regression and combined forecast model, respectively. In parenthesis are the results of the Pesaran and Timmermann (1992) directional accuracy test

freedom.[3] The forecasting ability of competing models was evaluated by using the likelihood ratio test of correct conditional coverage, which takes into account independence and unconditional coverage (see, e.g., Jorion 2007) for a significance level of 1%. Additionally, we evaluated the competing models through the VaR-based investor loss function described by Sarma et al. (2003) and given by $l_t^h = \left(r_t - VaR_t^h\right)^2 I_{\left\{r_t < VaR_t^h\right\}}$, where $I_{\{.\}}$ is the usual indicator function. Defining the loss differential between a model h with the RW model as $z_t = l_t^h - l_t^{RW}$, we can test the null of a zero-median loss differential against the alternative of a negative-median loss differential through the one-sided sign test given by $\hat{S}_{h,RW} = \left(S_{h,RW} - 0.5T\right)(0.25T)^{-0.5}$, which is asymptotically distributed as a standard normal, with $S_{h,RW} = \sum_{t=1}^{T} I_{\{Z_t \geq 0\}}$. Hence, since $\hat{S}_{RW,h} < -1.645$, the null can be rejected. Table 6 shows that the results for the correct conditional coverage are similar across high and low volatility periods. Although each model was rejected as not having the correct conditional coverage for the 5-min period, no model was rejected for the longer periods. Clear-cut evidence was found on using the one-side sign test: for all horizons and subsamples, the linear AR model and all the nonlinear specifications outperformed the simple RW model. When we compared nonlinear models with the linear AR model, the latter was improved by the nonlinear specifications. This evidence supports the usefulness of nonlinear modelling for risk management tasks.

[3] For the MS model, the VaR at time t is obtained as a weighted average of the VaR in each regime at time t, using as weights the probabilities of each state (see Billio and Pellizon 2000).

Table 6 Risk management out-of-sample VaR evaluation for all models and periods

	5 April 2006 to 3 August 2006				1 February 2011 to 13 May 2011			
	5 mins	10 mins	30 mins	60 mins	5 mins	10 mins	30 mins	60 mins
RW	0.03	0.29	0.65	0.73	0.04	0.47	0.77	0.41
AR	0.01	0.11	0.96	0.72	0.04	0.47	0.98	0.14
	(−36.32)	(−25.28)	(−14.56)	(−9.65)	(−33.23)	(−23.37)	(−13.57)	(−9.11)
MS	0.03	0.54	0.96	0.73	0.01	0.21	0.77	0.41
	(−36.26)	(−25.36)	(−14.56)	(−9.45)	(−33.29)	(−23.54)	(−13.57)	(−9.33)
ANN	0.01	0.11	0.96	0.73	0.04	0.47	0.99	0.97
	(−36.32)	(−25.28)	(−14.56)	(−9.65)	(−33.23)	(−23.37)	(−13.57)	(−8.90)
SVM	0.01	0.11	0.96	0.73	0.04	0.47	0.99	0.14
	(−36.32)	(−25.28)	(−14.56)	(−9.85)	(−33.23)	(−23.37)	(−13.57)	(−9.11)
COMB	0.03	0.54	0.96	0.59	0.11	0.74	0.77	0.77
	(−36.15)	(−25.28)	(−14.56)	(−9.05)	(−33.23)	(−23.37)	(−13.57)	(−8.47)

Note. RW, AR, MS, ANN, SVM and COMB denote random walk model, autoregressive model, Markov switching regression model, artificial neural network, support vector machine regression and combined forecast model, respectively. In parenthesis are the one-sided sign tests for competing model evaluation

4 Conclusions

We studied the intraday predictability of stock returns for different time horizons using linear and nonlinear models and taking into account fat tails and volatility dynamics. Using S&P 500 intraday data for two sample periods with different trend and volatility behaviour and within-day temporal horizons of 5, 10, 30 and 60 min, we compared the out-of-sample forecasting ability of a simple RW model and an AR regression model with nonlinear models—such as two-state MS regression, ANN, SVM regression and a combined forecast model—that potentially capture nonlinearity-in-mean in intraday stock returns. Statistical criteria like the mean squared forecast error and out-of-sample log-likelihood evaluations indicated somewhat limited evidence for predictability with respect to simple linear models, thus providing evidence in favour of intraday market efficiency. This evidence is robust across different sample periods and against data-snooping bias and the model overfitting problem; it is also consistent with results for high-frequency data reported in the literature (e.g., in Wang and Yang (2010) for energy markets).

Interestingly, economic criteria uncovered more robust evidence of predictability. Taking a simple trading rule driven by model forecasts, we found that the returns for this strategy based on the SVM regression forecast for a 60-min period yielded the best out-of-sample profitability for both high and low volatility sample periods. Likewise, in terms of risk evaluation, we found that nonlinear models were particularly beneficial for investors seeking to minimize the effects of unfavourable market fluctuations. Nonetheless, we were unable to find significant differences in market timing ability between linear and nonlinear models.

Acknowledgments We thank two anonymous referees and participants in the Forecasting Financial Markets Conference 2009, Computational and Financial Econometrics CFE09 and the 10th International Conference on Computational and Mathematical Methods in Science and Engineering CMMSE 2010 for constructive comments that greatly improved the previous version of this paper. Financial support from the MTM2008-03010 project is gratefully acknowledged. The first author acknowledges financial support provided by the Xunta de Galicia under research grant INCITE09201042PR.

References

Bali, T. G., Demirtas, K. O., & Levy, H. (2008). Nonlinear mean reversion in stock prices. *Journal of Banking and Finance, 32*, 767–782.
Bianco, S., & Renò, R. (2006). Dynamics of intraday serial correlation in the Italian futures market. *Journal of Futures Markets, 26*, 61–84.
Billio, M., & Pellizon, L. (2000). Value-at-Risk: A multivariate switching regime approach. *Journal of Empirical Finance, 7*(5), 531–554.
Bousquet, G., & Elisseeff, A. (2002). Stability and generalization. *Journal of Machine Learning Research, 2*, 499–526.
Brooks, C., & Persand, G. (2003). Volatility forecasting for risk management. *Journal of Forecasting, 22*, 1–22.
Campbell, J., Lo, A., & MacKinlay, C. (1997). *The econometrics of financial markets*. Princeton, NJ: Princeton University Press.
Cecchetti, S. G., Lam, P.-S., & Mark, N. C. (1990). Mean reversion in equilibrium asset prices. *American Economic Review, 80*, 398–418.
Chaudhuri, K., & Wu, Y. (2003). Random walk versus breaking trend in stock prices: Evidence from emerging markets. *Journal of Banking and Finance, 27*, 575–592.
Chen, X., & White, H. (1999). Improved rates and asymptotic normality for nonparametric neural networks estimators. *IEEE Transactions on Information Theory, 45*, 682–691.
Chew, L. (1994). Shock treatment. *Risk, 7*, 63–70.
Chordia, T., Roll, R., & Subrahmanyam, A. (2005). Evidence on the speed of convergence to market efficiency. *Journal of Financial Economics, 76*, 271–292.
Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal*.
Clements, M. P., & Taylor, N. (2003). Evaluating interval forecasts of high-frequency financial data. *Journal of Applied Econometrics, 18*, 445–456.
Dacco, R., & Satchell, S. (1999). Why do regime-switching models forecast so badly?. *Journal of Forecasting, 18*(1), 1–16.
Dueker, M. J. (1997). The econometrics of ultra-high frequency data. *Journal of Business and Economic Statistics, 15*(1), 26–34.
Fama, E. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance, 25*, 383–417.
Fama, E. (1991). Efficient capital markets: II. *Journal of Finance, 46*, 1575–1617.
Fine, T. (1999). *Feedforward neural network methodology*. New York: Springer.
Gencay, R. (1998). The predictability of security returns with simple technical trading rules. *Journal of Empirical Finance, 5*, 347–359.
Gencay, R. (1999). Linear, nonlinear and essential foreign exchange rate prediction with simple trading rules. *Journal of International Economics, 47*, 91–107.
Granger, C. W. J., & Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting, 3*, 197–204.
Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics, 42*, 27–62.
Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica, 57*, 357–384.
Hamilton, J. D. (1994). *Time series analysis*. Princeton, NJ: Princeton University Press.
Hastie, T., Tibshirani, R. J., & Friedman, J. (2001). *The elements of statistical learning. Data mining, inference, and prediction*. New York: Springer.
Haykin, S. (1999). *Neural networks. A comprehensive foundation*. Upper Saddle River: Prentice Hall.

Hong, Y., & Lee, T. H. (2003). Inference on predictability of foreign exchange rates via generalized spectrum and nonlinear time series models. *Review of Economics and Statistics, 85*, 1048–1062.
Hong, Y., Li, H., & Zhao, F. (2007). Can the random walk model be beaten in out-of-sample density forecasts?. Evidence from intraday foreign exchange rates. *Journal of Econometrics, 141*, 736–776.
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.
Hsiec, D. A. (1989). Testing for nonlinear dependence in daily foreing Exchange rates. *Journal of Business, 62*, 339–368.
Hsieh, D. A. (1991). Chaos and nonlinear dynamics: Application to financial markets. *Journal of Finance, 46*, 1839–1877.
Hsiec, D. A. (1993). Implications of nonlinear dynamics for financial risk management. *Journal of Financial and Quantitative Analysis, 28*, 41–64.
Hwang, J. T. G., & Ding, A. A. (1997). Prediction intervals for artificial neural networks. *Journal of the American Statistical Association, 92*, 748–757.
Jorion, P. (2007). *Value at Risk: The new benchmark for managing financial risk* (3rd ed.). New York: McGraw Hill.
Kahneman, D., & Tversky, A. (2000). *Choices, values, and frames*. Cambridge: Cambridge University Press.
Krzyzak, A., Linder, T., & Lugosi, G. (1996). Nonparametric estimation and classification using radial basis function nets and empirical risk minimization. *IEEE Transactions on Neural Networks, 7*, 475–487.
Lee, C. I., Gleason, K. C., & Mathur, I. (2000). Efficiency tests in the French derivative market. *Journal of Banking and Finance, 24*, 787–807.
Leitch, G., & Tanner, J. E. (1991). Economic forecast evaluation: Profit versus the conventional error measures. *American Economic Review, 81*, 580–590.
Leshno, M., Lin, V. Y., Pinkus, A., & Schocken, S. (1993). Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks, 6*, 861–867.
Liu, S. (2007). Currency derivatives and exchange rate forecastability. *Financial Analysts Journal, 63*(4), 72–78.
Liu, C. Y., & He, J. (1991). A variance ratio test of random walks in foreign exchange rates. *Journal of Finance, 46*, 773–785.
Lo, A. W. (2004). The adaptive market hypothesis: Market efficiency from an evolutionary perspective. *Journal of Portfolio Management, 30*, 15–29.
Lo, A. W., & Mackinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies, 1*, 41–66.
Mcqueen, G., & Thorley, S. (1991). Are stock returns predictable? A test using Markov chains. *Journal of Finance, 46*, 239–263.
Medeiros, M. C., Veiga, A., & Pedreira, C. E. (2001). Modelling exchange rates: Smooth transitions, neural networks, and linear models. *IEEE Transactions on Neural Networks, 12*(4), 755–764.
Meese, R. A., & Rose, A. K. (1990). Nonlinear, nonparametric, nonessential exchange rate estimation. *American Economic Review*, Papers and Proceedings, *80*, 192–196.
Meese, R. A., & Rose, A. K. (1991). An empirical assessment of nonlinearities in models of exchange rate determination. *Review of Economic Studies 58*, 603–619, of Forecasting, *5*, 559–583.
Pan, M., Chan, K. C., & Fok, R. C. W. (1997). Do currency futures prices follow random walks?. *Journal of Empirical Finance, 4*, 1–15.
Park, J., & Sandberg, I. W. (1993). Approximation and radial-basis-function networks. *Neural Computation, 5*, 305–316.
Patro, D. K., & Wu, Y. (2004). Predictability of short-horizon returns in international equity markets. *Journal of Empirical Finance, 11*, 553–584.
Pérez-Rodríguez, V., Torra, S., & Andrada-Félix, J. (2005). STAR and ANN models: Forecasting performance on the Spanish "IBEX-35" stock index. *Journal of Empirical Finance, 12*, 490–509.
Pesaran, M. H., & Timmermann, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business and Economic Statistics, 10*, 461–465.
Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in C*. Cambridge: Cambridge University Press.
Rapach, D. E., & Wohar, M. E. (2006). The out-of-sample forecasting performance of nonlinear models of real exchange rate behavior. *International Journal of Forecasting, 22*, 341–361.

Sarma, M. S., Thomas, S., & Shah, A. (2003). Selection of Value-at-Risk models. *Journal of Forecasting, 22*, 337–358.

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

Schölkopff, B., & Smola, A. J. (2002). *Learning with Kernels*. Cambridge: MIT Press.

Steinwart, I. (2002). Support vector machines are universally consistent. *Journal of Complexity, 18*, 768–791.

Swanson, N. R., & 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). A model selection approach to real time macroeconomic forecasting using linear models and artificial neural networks. *Review of Economics and Statistics, 79*, 540–550.

Timmermann, A. (2008). Elusive return predictability. *International Journal of Forecasting, 24*, 1–18.

Trapletti, A., Leisch, F., & Hornik, K. (2000). Stationarity and integrated autoregressive neural network processes. *Neural Computation, 12*, 2427–2450.

Vapnik, V. (1998). *Statistical learning theory*. New York: Wiley.

Wang, T., & Yang, J. (2010). Nonlinearity and intraday efficiency tests on energy futures markets. *Energy Economics, 32*, 496–503.

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

Yang, Y. (2004). Combining forecasting procedures: Some theoretical results. *Econometric Theory, 20*, 176–222.