INVESTIGACIONES ECONOMICAS. Volumen XXI (1) - Enero, 1997, pp. 75-91

USING NEAREST-NEIGHBOUR PREDICTORS TO FORECAST THE SPANISH STOCK MARKET*

Fernando FERNANDEZ-RODRIGUEZ
Universidad de Las Palmas de Gran Canaria
Simón SOSVILLA-RIVERO
FEDEA and Universidad Complutense de Madrid
María Dolores GARCIA-ARTILES
Universidad de Las Palmas de Gran Canaria

In this paper we apply local predictors, inspired by the literature on forecasting in nonlinear systems, to the General Index of the Madrid Stock Market for the period 2 January 1968-31 January 1994. When forecasting performance is measured by Theil's U statistic, our local predictors perform marginally better than a random walk, outperforming the random walk directional forecast. When formally testing for forecast accuracy, the results are mixed. Finally, when assessing the economic value of the local predictors, the results of using them as a filter technique are always superior to a buy-and-hold strategy assuming zero transaction costs.

1. Introduction

Forecasting financial markets has always been of great interest. Recent advances in both analytic and computational methods have helped empirical analysis of nonlinear models and have fuelled the research agenda in this area, dramatically increasing the number of approaches to forecasting financial time series. On the one hand, we have, among others, the ARCH and GARCH nonlinear stochastic processes introduced by Engle (1982) and Bollerslev (1986) (see Bollerslev, Chou and Kroner, 1992, for a review), the threshold autorregressive model of Tong and Lin (1980), the bilinear model of Granger and Andersen (1978), the switching regime model of Hamilton (1989) and the process of Clark (1973). On the other hand, we have the nonlinear deterministic models, where the studies of chaotic motions can be situated (see, e.g., references contained in Trippi, 1995).

* The authors wish to thank two anonymous referees and Rafael Repullo for very helpful comments, InterMoney for providing the data set and Julián Andrada-Félix for his computing assistance. Financial support by the Spanish Ministry of Education, through DGICYT Project PB94-0425 is also gratefully acknowledged.

Special attention has been paid to testing for predictable components in stock prices (see, e.g., Fama and French, 1988, and Lo and Mackinlay, 1988, and Fama, 1991, for a review). More recently, the existence of patterns in asset prices has been exploited to improve stock-market forecastability using different techniques: Artificial Neural Networks (ANNs) (see Van Eyden, 1995, for a review, and Olmeda and Pérez, 1995, for an analysis of the Spanish stock market), nonparametric analysis (nearest neighbours in Diebold and Nason, 1990, or Multivariate Adaptative Regression Splines (MARS) in Lewis, Ray and Stevens, 1994) or parametric analysis (such as the bilinear model in Weiss, 1986), to name but a few studies. Empirical studies tend to confirm some short-term predictability, as suggested by the technical analysis (see, e.g., Pring, 1985).

This paper examines whether nearest-neighbour prediction methods can improve out-of-sample forecasting in the Spanish stock market. To that end, we apply both the univariate local predictor presented in Farmer and Sidorowich (1987) an the multivariate case presented in Fernández-Rodríguez, Sosvilla-Rivero and Andrada-Félix (1996). Our local predictors, based on the literature on forecasting in nonlinear systems, allow us to make short-term forecasting of the future behaviour of a time series using information based on past values only.

The paper is organised as follows. The local predictors are presented in Section 2, while Section 3 describes the data set and offers some preliminary results. In Section 4 the forecast accuracy of the local predictors is assessed from a statistical point of view. In Section 5 we use the local predictions in a simple trading strategy in order to assess the economic significance of predictable patterns in the stock market. Some concluding remarks are provided in Section 6.

2. Prediction by nearest-neighbour methods

2.1. The univariate case

Let x_t $(t = 1,\ldots,n)$ be a finite time series. In order to detect behavioural patterns in this series, segments of equal length are considered as vectors $x_t^{m,\tau}$ of m observations sampled from the original time series at intervals of $\tau \in \mathbb{N}$ periods:

$$x_t^{m,\tau} = (x_t, x_{t-\tau}, \ldots, x_{t-(m-1)\tau}),\ t = 1 + \tau(m\text{-}1),\ldots,\ n \qquad [1]$$

with m referred to as the *embedding dimension* and τ called the *delay parameter.* These m-dimensional vectors are often called *m-histories,* while the m-dimensional space $\mathbb{R}^m$ is referred to as the phase space of time series.

The sequence of m-histories constitutes a m-dimensional object that can, for a big enough m, mimic the data generation process (Takens, 1981).

In order to simplify, we shall only consider the case of $\tau = 1$ and we shall write $x_t^{m,1} = x_t^{m1}$.

The proximity of two m-histories in the phase space $\mathbb{R}^m$ allows us to talk of «nearest neighbours» in the dynamic behaviour of two segments in the time series x_t.

In order to generate short-run forecasts, and as an alternative to the traditional Box-Jenkins methodology, we use the nearest-neighbour approach proposed by Farmer and Sidorowich (1987). This approach was previously applied to exchange rates in Bajo-Rubio, Fernández-Rodríguez and Sosvilla-Rivero (1992a,b). It is based on the analysis of time series in the state space domain and exploits spatial correlations to improve short term forecasts of nonlinear processes. The basic idea behind these predictors is that pieces of time series sometime in the past might have a resemblance to pieces in the future. In order to generate predictions, similar patterns of behaviour are located in terms of nearest neighbours. The time evolution of these nearest neighbours is exploited to yield the desired prediction. Therefore, the procedure only uses information local to the points to be predicted and makes no attempt to fit a function to the whole time series at once.

This aproach does not require stationarity in the data series x_t ($t = 1,\ldots,n$), the local predictions being generated by analysing the historical paths of the vectors around the last available vector.

$$x_n^m = (x_n,\ x_{n-1},\ x_{n-2},\ \ldots,\ x_{n-(m-1)}). \qquad [2]$$

Segments with similar dynamic behaviour are detected and used to produce the forecast, which is computed as some average of actually observed terms next to the segments involved. Therefore, in order to construct a local predictor we have considered the k m-histories

$$x_{i_1}^m,\ x_{i_2}^m,\ x_{i_3}^m,\ldots,\ x_{i_k}^m, \qquad [3]$$

most similar to x_n^m. The future short-term evolution of the time series will then be obtained using the information contained in the nearest neighbours found in the past.

In order to establish nearest neighbours to x_n^m, we look for the closest k vectors [3] in the phase space $\mathbb{R}^m$, in the sense that they maximise the function:

$$\rho(x_i^m,\ x_n^m) = \frac{cov\ (x_i^m,\ x_n^m)}{\sqrt{var\ (x_i^m)\ \text{var}\ (x_n^m)}}, \qquad [4]$$

[1] Setting the delay equal to one implies that vectors in the m-dimensional embedding space have an alternative representation as segments of m sequential observations from the original time series.

where *cov* and *var* denote covariance and variance, respectively. As it can be seen in [4], we look for the highest serial correlation of all m-histories, x_i^m, with the last one, x_n^m[2].

Once the nearest neighbours to x_n^m have been established, we consider predictors of the future evolution of x_n^m. A predictor is simply a rule for obtaining an estimate for the next observation x_{n+1}. The prediction $\hat{x}_{n+1}$ of x_{n+1} can be obtained using some extrapolation of the observations

$$x_{i_1+1},\ x_{i_2+1}\cdots,\ x_{i_k+1} \qquad [5]$$

subsequent to the k nearest neighbours m-histories that have been chosen, that is to say:

$$\hat{x}_{n+1} = F(x_{i_1+1},\ x_{i_2+1}\cdots,\ x_{i_k+1})$$

When generating nearest neighbours predictions, locally adjusted linear autoregressive predictions are usually employed:

$$\hat{x}_{n+1} = \hat{a}_0 x_n + \hat{a}_1 x_{n-1} + \ldots + \hat{a}_{m-1} x_{n-(m+1)} + \hat{a}_m \qquad [6]$$

whose coefficients have been fitted by ordinary least squares: that is, through a regression on the k nearest neighbours chosen [3] of their future evolution [5]. Sugihara and May (1990) and Casdagli and Weigend (1994) offer a detailed exposition of this kind of predictor.

2.2. The multivariate case

The above approach was extended to the multivariate case by Fernández-Rodríguez, Sosvilla-Rivero and Andrada-Félix (1996). To simplify notation, let us consider a set of two time series: x_t $(t = 1,\ldots,n)$ and y_t $(t = 1,\ldots,n)$.

We are interested in marking predictions of an observation of one of these series (e.g., x_{n+1}) by simultaneously considering nearest neighbours in both series. To that end, we embed each of these series in the vector space $\mathbb{R}^{2m}$, and consider the following vector:

$$(x_n^m, y_n^m) \in \mathbb{R}^{2m}$$

which gives us the last available m-history for each time series.

[2] Alternatively, we could have established nearest neighbours to x_n^m by looking for the closest k points x_i^m that minimise the functions $||x_i^m - x_n^m||$ or $1\text{-}cos\ (x_i^m,\ x_n^m)$ (i.e., looking, respectively, for the minimum distance or the lowest angle with x_n^m). (See Fernández-Rodríguez, 1992, for the relationship between the nearest neighbours obtained by using these three functions).

In order to establish nearest neighbours to the last m-histories (x_n^m, y_n^m), we can look for the closest k points that maximise the function:

$$\rho(x_i^m, x_n^m) + \rho(y_i^m, y_n^m), \; i = m, m+1, \ldots, n. \qquad [7]$$

In this way, we obtain a set of k simultaneous m-histories in both series:

$$(x_{i_1}^m, y_{i_1}^m), (x_{i_2}^m, y_{i_2}^m), \ldots, (x_{i_k}^m, y_{i_k}^m).$$

The predictions for x_{n+1} and y_{n+1} can be obtained from a linear autoregressive predictor with varying coefficients estimated by ordinary least squares:

$$\hat{x}_{n+1} = \hat{a}_0 x_n + \hat{a}_1 x_{n-1} + \ldots + \hat{a}_{m-1} x_{n-(m+1)} + \hat{a}_m, \qquad [8a]$$

$$\hat{y}_{n+1} = \hat{b}_0 y_n + \hat{b}_1 y_{n-1} + \ldots + \hat{b}_{m-1} y_{n-(m+1)} + \hat{b}_m. \qquad [8b]$$

The difference between this predictor and that presented in [6] is that now the nearest neighbours are established using criteria in which both series are used.

3. Data and preliminary results

In this paper we have used daily closing prices of the General Index of the Madrid Stock Market (IGM) and the Standard and Poors 500 Index of the New York Stock Market (SP500). Our sample period runs from 2 January 1968 through 31 January 1994 (5477 observations).

Table 1 provides summary statistics of the price levels and returns series. As can be seen, the price level series are positively skewed and strongly serially correlated. The Jarque-Bera (1980) test for joint normal kurtosis and skewness rejects the normality hypothesis.

To test for unit roots, we have used the non-parametric tests proposed by Phillips and Perron (1988), which are robust to heteroscedasticity, deviations from normality and various forms of serial correlation in the univariate representation of the variables under the unit root null hypothesis. In Table 2 the statistics are reported for the levels and first differences[3]. The test statistics are highly supportive of a single unit root in each of the series. This is the standard empirical finding for asset prices (see, e.g., Mills, 1993).

Given the high correlation between both time series (see Table 1), we have also tested for causality in levels. The F-statistics from the Granger-causality test shown in Table 3 suggest unidirectional causality running from SP500 to IGM

[3] Using the Newey and West (1994) procedure, we established a truncation lag equal to 9.

TABLE 1
Summary statistics

	Levels		First differences	
	IGM	SP500	IGM	SP500
Mean	150.71	215.22	0.0003	0.0003
Maximum	358.31	568.77	0.089	0.087
Minimum	36.60	62.28	−0.097	−0.229
Std. deviation	98.41	137.01	0.009	0.010
Skewness	0.49	0.81	0.05	−2.06
Kurtosis	1.60	2.31	11.59	51.91
Jarque-Bera	716.43	770.16	18,117.2	591,239.8
Autocorrelation				
1	1.000	0.999	0.330	0.104
2	0.999	0.998	0.092	−0.034
3	0.998	0.998	0.042	−0.007
4	0.997	0.997	0.043	−0.033
Correlation	0.89		0.07	

TABLE 2
Phillips-Perron test statistics

	$Z(t_{\tilde{\alpha}})$	$Z(t_{\alpha^*})$	$Z(t_{\hat{\alpha}})$
	(1)	(2)	(3)
	Levels		
IGM	−1.22	0.03	1.35
SP500	−1.38	1.32	1.88
	First differences		
DIGM	−19.84[a]	−19.08[a]	−19.07[a]
DSP500	−20.00[a]	−19.93[a]	−19.83[a]

(1), (2) and (3) denote the Phillips-Perron statistics with drift and trend, with drift, and without drift, respectively (see Perron, 1988).
[a] denotes significance at the 1% level, using Mackinnon's (1991) extended tabulations of critical values.

(F value 116.21, significant at the 1% level)[4]. This result opens the possibility of using standard lineal methods in forecasting to consider the linear reaction of the Spanish market to the American one. In this paper we explore the forecasting possibilities of all kind of reactions between those markets[5].

[4] Using Akaike (1974) information criterion, the number of lags chosen to compute the Granger-causality test was 4.

[5] Our prediction method allows us to make use of both linear and nonlinear reactions in order to establish nearest neighbours.

TABLE 3
Granger-causality tests

	F-statistic	P-value
SP500 does not Granger cause IGM	116.21	0.00
IGM does not Granger cause SP500	0.86	0.49

Before computing our local predictors, we have tested for the presence of nonlinear dependence in the series, since evidence of nonlinearity would support our approach to forecasting. To that end, we have made use of the well-known BDS test statistic (see Brock, Dechert, Scheinkman and LeBaron, 1996):

$$BDS\,(m,\,\varepsilon) = \frac{\sqrt{n}\,[C_m(\varepsilon) - (C_1(\varepsilon))^m]}{\sigma_m(\varepsilon)}, \qquad [10]$$

where $C_m(\varepsilon)$ represents the fraction of all m-histories in the series that are «close» to (within ε of) each other and, $\sigma_m(\varepsilon)$ is an estimate of the standard deviation. Brock, Dechert, Scheinkman and LeBaron (1996) show that, under the null hypothesis of independent and identical distribution *(iid)*, the BDS statistic is asymptotically $N(0,1)$.

As for the practical implementation of the BDS test, given the presence of a unit root in the series, they were first-differenced. After that, we used the residual of the *AR (p)* model as inputs in order to remove any linear dependence in the time series[6]. On the other hand, since the *BDS* test statistic depends on the values of the embedding dimension and the chosen distance related to the standard deviation of the data (m and ε, respectively), following Hsieh (1989) and Brock, Hsieh and LeBaron (1991) in Table 4 we show the results for values of m from 2 and 7, and for values of ε ranging from 0.5σ to 2σ, where σ denotes the standard deviation of the series[7]. With over 5000 observations, we can use the normal standard tables to assess significance, since the small sample properties only become important for sizes less than 500 (see Brock, Hsieh and LeBaron, 1991). If the one percent marginal significant level is used, *iid* is rejected in 40 of the 42 statistics for IGM and in 29 of the 42 for SP500[8]. These results are in line with those of Ramsey (1990), Hsieh (1991), Pununzi and Ricci (1993), and Olmeda and Pérez (1995), among others. This opens alternatives of nonlinear dependence as well as nonstationarity. Although nonstationarity of the series is detected, in this paper we explore the use of nonlinear dependencies in order to forescast the series[9].

[6] Using the Schwartz information criterion, the appropriate lag length p was set equal to 4.

[7] The statistics were calculated using the *BDS-STATS* program written by W. Dechert.

[8] If we use the 5 percent marginal significant level, *iid* is rejected in 41 of the 42 statistics for IGM and in 33 of the 42 for SP500.

[9] Note also that ARCH effects could cause the rejection of the BDS, but they do not imply forecastability.

TABLE 4
BDS test statistics

	(A) IGM residuals						
	$\varepsilon=0.5\sigma$	$\varepsilon=0.75\sigma$	$\varepsilon=\sigma$	$\varepsilon=1.25\sigma$	$\varepsilon=1.5\sigma$	$\varepsilon=1.75\sigma$	$\varepsilon=2\sigma$
$m = 2$	6.00[a]	5.88[a]	4.89[a]	3.83[a]	2.93[a]	2.23[b]	1.71[c]
$m = 3$	9.44[a]	10.75[a]	9.83[a]	8.25[a]	6.61[a]	5.19[a]	4.07[a]
$m = 4$	10.72[a]	13.84[a]	13.76[a]	12.26[a]	10.34[a]	8.47[a]	6.55[a]
$m = 5$	10.78[a]	15.52[a]	16.52[a]	15.47[a]	13.59[a]	11.53[a]	9.55[a]
$m = 6$	10.35[a]	16.29[a]	18.39[a]	18.05[a]	16.43[a]	14.39[a]	12.20[a]
$m = 7$	9.75[a]	16.49[a]	19.53[a]	19.90[a]	18.67[a]	16.82[a]	14.57[a]
	(B) SP500 residuals						
	$\varepsilon=0.5\sigma$	$\varepsilon=0.75\sigma$	$\varepsilon=\sigma$	$\varepsilon=1.25\sigma$	$\varepsilon=1.5\sigma$	$\varepsilon=1.75\sigma$	$\varepsilon=2\sigma$
$m = 2$	1.53	2.03[b]	2.05[b]	1.78[c]	1.42	1.08	0.82
$m = 3$	2.02[b]	3.54[a]	4.16[a]	3.96[a]	3.36[a]	2.69[a]	2.11[b]
$m = 4$	1.93[c]	4.37[a]	5.87[a]	6.06[a]	5.45[a]	4.55[a]	3.67[a]
$m = 5$	1.62	4.62[a]	7.08[a]	7.95[a]	7.55[a]	6.52[a]	5.36[a]
$m = 6$	1.25	4.46[a]	7.74[a]	9.40[a]	9.38[a]	8.35[a]	7.01[a]
$m = 7$	0.93	4.08[a]	7.96[a]	10.43[a]	10.92[a]	10.05[a]	8.63[a]

The *BDS* statistic is applied to the ARIMA (4, 1, 0) residuals of the original series.
[a], [b] and [c] denote significance at the 1%, 5% and 10% levels, respectively.
σ is the standard deviation of the series.

Scheinkman and LeBaron (1989) propose to recreate the data series by sampling randomly with replacement from the data until one has a «shuffled» series of the same length as the original. The shuffled series should be completely random (though preserving the original distribution). Applying the BDS test to the shuffled residuals series, the null hypothesis of *iid* is retained, because all BDS test values are less than the critical values (see Table 5). Therefore, there is evidence that some nonlinear structure present in the original series has been removed by shuffling[10].

Based on the indications of nonlinearities previously reported, we proceeded to assess the forecasting performance of our predictors for the IGM in both versions: univariate and multivariate. In the latter case, given the results of the Granger-causality tests, the SP500 series is used for establishing nearest neighbours.

Since our predictors depend on the values of embedding dimension m and the number of closest k points in the phase space $\mathbb{R}^m$, we present the results for values of m between 2 and 7 and for values of k of 90, 100, 110, 120, 130 and 140. We used for modelling the first 5222 observations, keeping the last 255 to test the forecast accuracy. Our local predictors were used to produce forecasts one day ahead from the observation 5223. Then, the data

[10] These results validate the size ot the test under the null (Brock, Hsieh and LeBaron, 1991).

TABLE 5
BDS test statistics

	(A) Shuffled IGM residuals						
	$\varepsilon=0.5\sigma$	$\varepsilon=0.75\sigma$	$\varepsilon=\sigma$	$\varepsilon=1.25\sigma$	$\varepsilon=1.5\sigma$	$\varepsilon=1.75\sigma$	$\varepsilon=2\sigma$
$m = 2$	0.16	0.17	0.10	0.08	0.09	0.10	0.08
$m = 3$	0.08	0.13	0.10	0.12	0.12	0.12	0.08
$m = 4$	−0.02	0.04	0.05	0.05	0.05	0.04	−0.02
$m = 5$	−0.05	−0.01	0.02	−0.01	−0.02	0.01	−0.05
$m = 6$	−0.05	−0.04	−0.03	−0.08	−0.09	−0.02	−0.07
$m = 7$	−0.02	−0.01	−0.01	−0.07	−0.08	0.03	−0.02
	(B) Shuffled SP500 residuals						
	$\varepsilon=0.5\sigma$	$\varepsilon=0.75\sigma$	$\varepsilon=\sigma$	$\varepsilon=1.25\sigma$	$\varepsilon=1.5\sigma$	$\varepsilon=1.75\sigma$	$\varepsilon=2\sigma$
$m = 2$	−0.14	−0.17	−0.16	−0.11	−0.05	−0.01	0.01
$m = 3$	−0.13	−0.23	−0.27	−0.23	−0.14	−0.07	−0.03
$m = 4$	−0.04	−0.09	−0.14	−0.12	−0.04	0.01	0.03
$m = 5$	−0.01	−0.01	−0.04	−0.04	−0.03	0.08	0.10
$m = 6$	0.01	0.02	0.01	0.01	0.08	0.15	0.17
$m = 7$	0.01	0.04	0.06	0.08	0.17	0.25	0.29

The *BDS* statistic is applied to the shuffled ARIMA (4, 1, 0) residuals of the original series.
σ is the standard deviation of the series.
The critical values are 2.58, 1.96 and 1.64 for the 1%, 5% and 10% significance levels, respectively.

for this observation was added to the sample, the models were re-estimated, and new forecasts were generated for the IGM time series. This recursive process continued until forecasts were generated using the observation 5476.

The forecasting performance was initially measured by Theil's U statistic, defined as the ratio of the root mean square error (RMSE) of forecast from our predictors to the RMSE of the naive random walk forecast. Therefore, a value of U less than one indicates better performance than the random walk specification.

As can be seen in Table 6 (upper part), the U statistics are, for the univariate case, greater than one only in two of the 36 cases, and less than 0.99 in 30 of the 36 cases, suggesting that our predictors marginally outperform the random walk. Note that our best univariate local predictor presents an improvement of 2.98%.

From the lower part of Table 6, we see that in all 36 cases, the multivariate predictors offer lower U statistics than the univariate case, the best multivariate local predictor showing an improvement of 5.95% out-of-sample.

A further test of forecasting performance relative to the forecasts of a random walk is the accuracy in predicting the direction of IGM movements. To that end, we have also computed the percentage of correct predictions.

TABLE 6
Theil's U statistic

	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$	$m = 7$
(A) Univariate case						
$k = 90$	0.9898	1.0004	0.9796	0.9771	0.9705	0.9786
$k = 100$	0.9894	1.0004	0.9771	0.9753	0.9739	0.9706
$k = 110$	0.9860	0.9989	0.9799	0.9788	0.9760	0.9706
$k = 120$	0.9893	0.9949	0.9801	0.9791	0.9757	0.9702
$k = 130$	0.9883	0.9946	0.9807	0.9727	0.9730	0.9734
$k = 140$	0.9862	0.9950	0.9824	0.9727	0.9763	0.9701
(B) Multivariate case						
	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$	$m = 7$
$k = 90$	0.9725	0.9786	0.9699	0.9750	0.9441	0.9620
$k = 100$	0.9709	0.9716	0.9638	0.9722	0.9405	0.9576
$k = 110$	0.9692	0.9719	0.9647	0.9693	0.9431	0.9585
$k = 120$	0.9694	0.9683	0.9616	0.9689	0.9446	0.9570
$k = 130$	0.9729	0.9664	0.9610	0.9668	0.9479	0.9545
$k = 140$	0.9689	0.9675	0.9599	0.9649	0.9482	0.9587

Our local predictors always show a value higher than 0.5, clearly outperforming the random walk directional forecast (see Table 7)[11]. Note also that in 30 out of the 36 cases, the multivariate predictors offer higher values than the univariate case.

TABLE 7
Directional forecast

	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$	$m = 7$
(A) Univariate case						
$k = 90$	0.5412	0.5529	0.5608	0.5765	0.5843	0.5922
$k = 100$	0.5529	0.5490	0.5843	0.5765	0.5726	0.5961
$k = 110$	0.5647	0.5176	0.5608	0.5804	0.5686	0.5882
$k = 120$	0.5642	0.5294	0.5882	0.5804	0.5726	0.5804
$k = 130$	0.5725	0.5294	0.5804	0.5922	0.5647	0.5922
$k = 140$	0.5745	0.5490	0.5961	0.5726	0.5647	0.5922
(B) Multivariate case						
	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$	$m = 7$
$k = 90$	0.6118	0.5961	0.5882	0.5843	0.6118	0.5882
$k = 100$	0.6118	0.6000	0.5882	0.5765	0.6196	0.5922
$k = 110$	0.5961	0.5882	0.6039	0.5765	0.6118	0.6039
$k = 120$	0.6159	0.5922	0.5922	0.5804	0.6157	0.5961
$k = 130$	0.6000	0.6000	0.6000	0.5922	0.6118	0.5961
$k = 140$	0.6118	0.6157	0.6000	0.5961	0.5922	0.6039

Percentage of correct forecast direction.

[11] The value 0.5 is the usual benchmark. However, the results must be treated with caution, since numbers of positive changes do not necessarily coincide with the number of negative changes.

Therefore, the evidence presented in Tables 6 and 7 suggests that in predicting the IGM time series some forecast accuracy can be gained by considering the information content of other related stock prices (in our case, SP500). The next step is to test for the statistical significance of the differences in forecast accuracy.

4. Assessing forecast accuracy of the local predictors

We have considered two tests of forecast accuracy: the Williams-Kloot test described in Williams (1959) and the test recently proposed by Diebold and Mariano (1995). Let f_1 and f_2 denote alternative forecasts of the variable z, let e_1 and e_2 denote the corresponding forecast errors ($z - f_1$ and $z - f_2$, respectively), and let $d = e_1^2 - e_2^2$ denotes the loss differential. The Williams-Kloot test statistic is the t-ratio for the hypothesis that the coefficient on $f_1 - f_2$ is zero in a regression of $z - (f_1 + f_2)/2$ on $f_1 - f_2$. A significantly negative value implies that f_2 is statistically superior to that of f_1 (and *vice versa*). The Diebold-Mariano test involves a test of the hypothesis that the mean loss differential d is zero with an appropriate correction for serial correlation in the d series:

$$S = \bar{d} \left[\frac{2\pi \hat{f}_d(0)}{T} \right]^{-(1/2)}, \qquad [11]$$

where $\hat{f}_d(0)$ is a consistent estimate of the spectral density of the loss differential at frequency 0, T is the number of forecasts (255) and S is asymptotically distributed N (0,1).

The results are shown in Tables 8 and 9. As one can see, from the results of the Diebold-Mariano test, we do not reject the hypotheses of equal expected squared error (i. e., both the univariate and multivariate local predictors are not statistically significantly better predictors than the random walk). However, the Williams-Kloot test suggests that the multivariate predictor is superior (at least at 5% level) to the random walk in 16 of the 36 cases considered. Note also that in 7 out of the 36 cases, the multivariate predictors are superior to the univariate case[12].

5. Assessing the economic value of the local predictors

In order to assess the economic significance of predictable patterns in the IGM series, it is necessary to explicitly consider how investors may exploit the computed local predictions as a trading rule.

[12] If we use the 10 percent marginal significant level, the Williams-Kloot test suggests that the multivariate predictor is superior to the random walk in 21 of the 36 cases considered, while the multivariate predictors are superior to the univariate case in 13 out of the 36 cases.

TABLE 8
The Diebold-Mariano test statistic

(A) Random walk *vs.* univariate local predictor						
	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$	$m = 7$
$k = 90$	0.074	−0.001	0.253	0.402	0.535	0.451
$k = 100$	0.104	0.032	0.217	0.426	0.518	0.646
$k = 110$	0.225	0.064	0.243	0.392	0.505	0.576
$k = 120$	0.302	0.116	0.247	0.397	0.495	0.549
$k = 130$	0.313	0.088	0.316	0.530	0.539	0.486
$k = 140$	0.341	0.090	0.275	0.529	0.524	0.598
(B) Univariate *vs.* multivariate local predictors						
	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$	$m = 7$
$k = 90$	0.533	0.740	0.535	−0.024	0.510	0.129
$k = 100$	0.597	0.876	0.703	0.806	0.511	0.019
$k = 110$	0.472	0.827	0.686	0.254	0.529	0.205
$k = 120$	0.429	0.808	0.755	0.277	0.663	0.295
$k = 130$	0.312	0.860	0.586	0.068	0.469	0.518
$k = 140$	0.334	0.845	0.619	0.133	0.486	0.220
(C) Random walk *vs.* multivariate local predictor						
	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$	$m = 7$
$k = 90$	0.458	0.526	0.601	0.426	0.937	0.578
$k = 100$	0.511	0.658	0.715	0.532	0.936	0.700
$k = 110$	0.531	0.626	0.714	0.628	0.939	0.758
$k = 120$	0.586	0.663	0.761	0.645	1.006	0.799
$k = 130$	0.509	0.636	0.725	0.638	0.905	0.853
$k = 140$	0.543	0.624	0.715	0.678	0.893	0.784

TABLE 9
The Williams-Kloot test statistic

(A) Random walk *vs.* univariate local predictor						
	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$	$m = 7$
$k = 90$	−0.434	0.017	−0.987	−1.079	−1.481	−1.074
$k = 100$	−0.457	0.017	−1.118	−1.202	−1.315	−1.550
$k = 110$	−0.606	−0.047	−1.000	−1.056	−1.235	−1.578
$k = 120$	−0.465	−0.223	−1.002	−1.064	−1.268	−1.606
$k = 130$	−0.514	−0.239	−0.981	−1.417	−1.436	−1.462
$k = 140$	−0.614	−0.219	−0.889	−1.444	−1.299	−1.666[c]

TABLE 9 (Continued)
The Williams-Kloot test statistic

	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$	$m = 7$
(B) Univariate *vs.* multivariate local predictors						
$k = 90$	−1.034	−1.363	−0.672	−1.049	−1.884[c]	−1.262
$k = 100$	−1.198	−1.889[c]	−0.958	−0.212	−2.484[b]	−1.061
$k = 110$	−1.111	−1.884[c]	−1.139	−0.667	−2.606[a]	−1.030
$k = 120$	−1.320	−1.899[c]	−1.430	−0.725	−2.662[a]	−1.167
$k = 130$	−1.054	−2.143[b]	−1.561	−0.439	−2.213[b]	−1.808[c]
$k = 140$	−1.243	−2.166[b]	−1.810[c]	−0.621	−2.599[a]	−1.133
(C) Random walk *vs.* multivariate local predictor						
	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$	$m = 7$
$k = 90$	−1.168	−0.938	−1.466	−1.280	−3.001[a]	−2.127[c]
$k = 100$	−1.259	−1.270	−1.808[c]	−1.469	−3.272[a]	−2.448[c]
$k = 110$	−1.327	−1.273	−1.772[c]	−1.666[c]	−3.235[a]	−2.443[c]
$k = 120$	−1.323	−1.445	−1.970[b]	−1.710[c]	−3.251[a]	−2.573[c]
$k = 130$	−1.172	−1.554	−2.049[b]	−1.886[c]	−3.084[a]	−2.752[a]
$k = 140$	−1.350	−1.538	−2.142[b]	−2.021[b]	−3.131[a]	−2.536[b]

[a],[b] and [c] denote significance at the 1%, 5% and 10% levels respectively.

In this paper we combine a simple and widely popular trading strategy known as a filter technique, originally analysed by Alexander (1961) and Fama and Blume (1966), with the nonparametric forecasts and compare the return obtained with a buy and hold strategy. In this context, a δ% filter is defined as follows: If the daily closing «price» of the index is predicted to move up at least δ%, buy and hold until the price is predicted to move down at least δ% from a subsequent high, then simultaneously sell and go short. The short position is maintained until the daily closing price is predicted to rise at least δ% above a subsequent low when one covers and buys. Moves less than δ% in either direction are ignored.

We have computed the filter technique for 110 filters ranging in size from 0.05% to 5%. Given that in a random-walk market no mechanical trading rule would consistently outperform a buy-and-hold policy, we compare both strategies. Instead of showing the rate of return for each filter size, we have computed its average value, and then divided it by the average rate of return arising from a buy-and-hold strategy. Table 10 shows the results.

As can be seen in Table 10 (upper part), for the univariate case, the filter technique is always superior to a buy-and-hold strategy. In 25 of the 36 cases the filter technique improves the average returns by 2%, and in 10 of the 36 cases by 3%. In the best case, the improvement amounts to 3.47%.

From the lower part of Table 10, we see that in 29 out of the 36 cases, the multivariate predictors offer greater relative returns than the univariate case, the best case showing an improvement of 3.57% from the buy-and-hold strategy.

TABLE 10
Economic value of the forecasts

(A) Univariate case						
	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$	$m = 7$
$k = 90$	1.0269	1.0347	1.0306	1.0215	1.0198	1.0141
$k = 100$	1.0264	1.0345	1.0312	1.0224	1.0203	1.0176
$k = 110$	1.0262	1.0346	1.0306	1.0217	1.0195	1.0179
$k = 120$	1.0170	1.0337	1.0297	1.0203	1.0262	1.0206
$k = 130$	1.0246	1.0326	1.0300	1.0187	1.0186	1.0136
$k = 140$	1.0241	1.0300	1.0294	1.0181	1.0203	1.0125
(B) Multivariate case						
	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$	$m = 7$
$k = 90$	1.0354	1.0348	1.0268	1.0194	1.0260	1.0141
$k = 100$	1.0271	1.0357	1.0220	1.0216	1.0264	1.0228
$k = 110$	1.0344	1.0338	1.0279	1.0248	1.0248	1.0234
$k = 120$	1.0349	1.0347	1.0281	1.0251	1.0286	1.0244
$k = 130$	1.0352	1.0336	1.0304	1.0246	1.0257	1.0240
$k = 140$	1.0329	1.0314	1.0307	1.0222	1.0249	1.0235

Ratio between the average rate of the returns from the filtering technique and from the buy-and-hold policy.

Thus, assuming zero trading costs, the simple trading rule based on nonlinear predictions dominates the buy-and-hold strategy. However, to see if an investor who had followed nonlinear predictions as a trading rule could make profits net of transaction costs in excess of the payoffs from a passive buy-and-hold strategy, we would have to consider both the frequency at which transactions occur and the cost of trading, information that we do not have.

6. Concluding remarks

The purpose of out paper has been to contribute to the debate on the relevance of nonlinear forecasts of high-frequency data in financial markets. To that end, we have presented the first results of applying the local predictors introduced by Farmer and Sidorowich (1987) and in Fernández-Rodríguez, Sosvilla-Rivero and Andrada-Félix (1996) to the General Index of the Madrid Stock Market (IGM) for the period 2 January 1968-31 January 1994.

The main results are as follows. Firstly, when forecasting performance was measured by Theil's U statistic, the local predictors performed marginally better than a random walk in forecasting the IGM time series. Our best univariate local predictor presented an improvement of 2.98%, while our best multivariate local predictor showed an improvement of 5.95% out-of-sample. Furthermore, our local predictors outperformed the random walk when producing directional forecasts.

Secondly, we considered two tests of forecast accuracy. The Diebold-Mariano test suggested that both the univariate and multivariate local predictors were not statistically significantly better predictors than the random walk. On the other hand, the Williams-Kloot test showed that the multivariate predictor was superior to the random walk in 16 out of 36 cases, and that in only 7 out of 36 cases the multivariate predictors were superior to the univariate case.

Finally, when assessing the economic value of the local predictors, we found that the results of using nonlinear predictions in a filter technique were always superior to a buy-and-hold strategy. However, with transaction costs it could be possible that the return of an active strategy (many trades), such as the one proposed, could be inferior to a buy and hold strategy.

To conclude, note that the period covered is very long and heterogeneous, with a number of important events (e.g., oil crisis, changes in the computation of the index, increased number of trading days, the «crash» of 1987, the inception of the IBEX35 futures market, etc.) that could affect the structure of the series. A natural extension of this paper would be to examine the forecast accuracy of our predictors in different subperiods. Another possible extension would be to compare the forecast accuracy of our predictors related to the predictions of linear models that take into account the detected linear reaction of the Spanish market to the American one.

References

Akaike, H. (1974): «A new look at the statistical model identification», *IEEE Transactions of Automatic Control* 19, pp. 716-723.

Alexander, S. S. (1961): «Price movements in speculative markets: Trends or random walks?», *Industrial Management Review* 2, pp. 7-26.

Bajo-Rubio, O.; Fernández-Rodríguez, F. and Sosvilla-Rivero, S. (1992a): «Chaotic behaviour in exchange-rate series: First results for the peseta-US dollar case», *Economics Letters* 39, pp. 207-211.

Bajo-Rubio, O.; Fernández-Rodríguez, F. and Sosvilla-Rivero, S. (1992b): «Volatilidad y predecibilidad en las series del tipo de cambio peseta-dólar: Un enfoque basado en el caos determinista», *Revista Española de Economía*, Monográfico Mercados Financieros Españoles, pp. 91-109.

Bollerslev, Y. (1986): «Generalized autoregressive conditional heteroskedasticity», *Journal of Econometrics* 31, pp. 307-327.

Bollerslev, T.; Chou, R. Y. and Kroner, K. F. (1992): «ARCH modeling in finance: A review of the theory and empirical evidence», *Journal of Econometrics* 52, pp. 5-59.

Brock, W.; Hsieh, D. A. and LeBaron, B. (1991): *Nonlinear Dynamics, Chaos and Instability* (MIT Press, Cambridge, MA).

Brock, W.; Dechert, W.; Scheinkman, J. and LeBaron, B. (1996): «A test for independence based on the correlation dimension», *Econometric Reviews* 15, pp. 197-235.

Casdagli, M. and Weigend, A. S. (1994): «Exploring the continuum between deterministic and stochastic modelling», in A. S. Weigend and N. A. Gershenfeld (eds.) *Time Series Prediction: Forecasting the Future and Understanding the Past* (Reading, MA.: Addison Wesley).

Clark, P. K. (1973): «A subordinated stochastic process model with fine variance for speculative prices», *Econometrica* 41, pp. 135-155.

Diebold, F. X. and Mariano, R. S. (1995): «Comparing predictive accuracy», *Journal of Business and Economic Statistics* 13, pp. 253-263.

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

Engle, R. F. (1982): «Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation», *Econometrica* 50, pp. 987-1008.

Fama, E. F. (1991): «Efficient capital markets: II», *Journal of Finance* 46, pp. 1575-1617.

Fama, E. F. and Blume, M. E. (1966): «Filter rules and stock market trading», *Journal of Business* 39, pp. 226-241.

Fama, E. F. and French, K. R. (1988): «Dividend yields and expected stock returns», *Journal of Financial Economics* 22, pp. 3-25.

Farmer, D. and Sidorowich, J. (1987): «Predicting chaotic time series», *Physical Review Letters* 59, pp. 845-848.

Fernández-Rodríguez, F. (1992): *El problema de la predicción en series temporales: Aplicaciones del caos determinista,* Ph. D. Thesis, Universidad de Las Palmas de Gran Canaria.

Fernández-Rodríguez, F.; Sosvilla-Rivero, S. and Andrada-Félix, J. (1996): «Combining information in exchange rate forecasting: Evidence from the EMS», forthcoming in *Applied Economics Letters.*

Granger, C. W. J. and Andersen, A. P. (1978): *Introduction to Bilinear Models* (Gottien: Vandenhoek and Rupretcht).

Hamilton, J. D. (1989): «A new approach to the economic analysis of nonstationary time series and the business cycles», *Econometrica* 57, pp. 357-384.

Hsieh, D. A. (1989): «Testing for nonlinear dependence in daily foreign exchange rates», *Journal of Business* 62, pp. 339-368.

Hsieh, D. A. (1991): «Chaos and nonlinear dynamics: Application to financial markets», *Journal of Finance* 46, pp. 1839-1877.

Jarque, C. M. and Bera, A. K. (1980): «Efficient test for normality, homoscedasticity and serial independence of regression residuals», *Economics Letters* 6, pp. 255-259.

Lewis, P. A. W.; Ray, B. K. and Stevens, J. G. (1994): «Modeling time series using multivariate adaptative regression splines (MARS)», in A. S. Weigend and N. A. Gershenfeld (eds.), *Time Series Prediction: Forecasting the Future and Understanding the Past* (Reading, MA.: Addison Wesley).

Lo, A. W. and Mackinlay, A. C. (1988): «Stock prices do not follow random walks: Evidence from a simple specification test», *Review of Financial Studies* 1, pp. 41-66.

MacKinnon, J. G. (1991): «Critical values for cointegration tests», in R. F. Engle and C. W. J. Granger (eds.), *Long-run Economic Relationships* (Oxford: Oxford University Press).

Mills, T. C. (1993): *The Econometric Modelling of Financial Time Series* (Cambridge: Cambridge University Press).

Newey, W. K. and West, K. D. (1994): «Automatic lag selection in covariance matrix estimation», *Review of Economic Studies* 61, pp. 631-653.
Olmeda, I. and Pérez, J. (1995): «Non-linear dynamics and chaos in the Spanish stock market», *Investigaciones Económicas* 19, pp. 217-248.
Perron, P. (1988): «Trends and random walks in macroeconomic time series», *Journal of Economic Dynamics and Control* 12, pp. 297-332.
Phillips, P. C. B. and Perron, P. (1988): «Testing for unit root in time series regression», *Biometrika* 75, pp. 335-346.
Pring, J. M. (1985): *Technical Analysis Explained* (New York: McGraw Hill), 2nd. edition.
Pununzi, F. and Ricci, N. (1993): «Testing non linearities in Italian stock exchange», *Rivista Internazionale di Science Economiche e Commerciali* 40, pp. 559-574.
Ramsey, J. B. (1990): «Economic and financial data as nonlinear processes», in G. P. Dwyer and R. W. Hafer (eds.), *The Stock Market: Bubbles, Volatility, and Chaos* (Boston, MA.: Kluwer Academic), pp. 81-134.
Scheinkman, J. and LeBaron, B. (1989): «Nonlinear dynamics and stock returns», *Journal of Business* 62, pp. 311-337.
Sugihara, G. and May, R. M. (1990): «Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series», *Nature* 344, pp. 734-741.
Tong, H. and Lin, K. S. (1980): «Threshold autorregression, limit cycles and cyclical data», *Journal of the Royal Statistical Society, Ser. B* 42, pp. 245-292.
Trippi, R. (ed.) (1995): *Chaos and Nonlinear Dynamics in Financial Markets* (Chicago: Irwin).
Van Eyden, R. J. (1995): *The Application of Neural Networks in the Forecasting of Share Prices* (Haymarket, VA: Finance & Technology Publishing).
Weiss, A. A. (1986): «ARCH and bilinear time series models», *Journal of Business and Economic Statistics* 4, pp. 59-70.
Williams, E. J. (1959): *Regression Analysis* (Wiley, New York).

Resumen

Este trabajo aplica predictores locales, inspirados en la literatura de predicción en sistemas no lineales, al Indice General de la Bolsa de Madrid para el período enero de 1968-enero de 1994. Cuando medimos la bondad predictiva en términos del estadístico U de Theil, nuestros predictores se comportan marginalmente mejor que el paseo aleatorio, siendo muy superiores en predicción direccional. Al contrastar formalmente la bondad predictiva no obtenemos resultados concluyentes. Finalmente, examinamos su valor económico utilizándolos como regla técnica para comprar y vender, resultando siempre superiores, en ausencia de costes de transacción, a una estrategia de comprar y mantener.