RECOGNIZING AND FORECASTING THE SIGN OF FINANCIAL LOCAL TRENDS USING HIDDEN MARKOV MODELS

Manuele Bicego

Enrico Grosso

Edoardo Otranto

WORKING PAPERS

2008/03

CENTRO RICERCHE ECONOMICHE NORD SUD
(CRENoS)
UNIVERSITÀ DI CAGLIARI
UNIVERSITÀ DI SASSARI

Il CRENoS è un centro di ricerca istituito nel 1993 che fa capo alle Università di Cagliari e Sassari ed è attualmente diretto da Raffaele Paci. Il CRENoS si propone di contribuire a migliorare le conoscenze sul divario economico tra aree integrate e di fornire utili indicazioni di intervento. Particolare attenzione è dedicata al ruolo svolto dalle istituzioni, dal progresso tecnologico e dalla diffusione dell'innovazione nel processo di convergenza o divergenza tra aree economiche. Il CRENoS si propone inoltre di studiare la compatibilità fra tali processi e la salvaguardia delle risorse ambientali, sia globali sia locali.
Per svolgere la sua attività di ricerca, il CRENoS collabora con centri di ricerca e università nazionali ed internazionali; è attivo nell'organizzare conferenze ad alto contenuto scientifico, seminari e altre attività di natura formativa; tiene aggiornate una serie di banche dati e ha una sua collana di pubblicazioni.

www.crenos.it
info@crenos.it

CRENOS – CAGLIARI
VIA SAN GIORGIO 12, I-09100 CAGLIARI, ITALIA
TEL. +39-070-6756406; FAX +39-070- 6756402

CRENOS - SASSARI
VIA TORRE TONDA 34, I-07100 SASSARI, ITALIA
TEL. +39-079-2017301; FAX +39-079-2017312

Titolo: RECOGNIZING AND FORECASTING THE SIGN OF FINANCIAL LOCAL TRENDS USING HIDDEN MARKOV MODELS

Recognizing and forecasting the sign of financial local trends using Hidden Markov Models

Manuele Bicego[*], Enrico Grosso[†] and Edoardo Otranto[‡]

D.E.I.R., University of Sassari

Via Torre Tonda 34 - 07100 - Sassari - Italy.

Abstract

The problem of forecasting financial time series has received great attention in the past, from both Econometrics and Pattern Recognition researchers. In this context, most of the efforts were spent to represent and model the volatility of the financial indicators in long time series. In this paper a different problem is faced, the prediction of increases and decreases in *short* (local) financial trends. This problem, poorly considered by the researchers, needs specific models, able to capture the movement in the short time and the asymmetries between increase and decrease periods. The methodology presented in this paper explicitly considers both aspects, encoding the financial returns in binary values (representing the signs of the returns), which are subsequently modelled using *two* separate Hidden Markov models, one for increases and one for decreases, respectively. The approach has been tested with different experiments with the Dow Jones index and other shares of the same market of different risk, with encouraging results.

Keywords: Markov Models; Asymmetries; Binary data; Short-time forecasts.

JEL Codes: C02; C63; G11.

[*]University of Sassari, E-mail: bicego@uniss.it

[†]University of Sassari, E-mail: grosso@uniss.it

[‡]University of Sassari and CRENoS, E-mail: eotranto@uniss.it

1 Introduction

The increasing relevance of the financial markets in economy and the possibility of speculation on the Stock Exchange has favored the development of statistical and econometric models for financial markets. Great interest has been demonstrated by researchers of both econometrics and pattern recognition research areas, with many methodologies proposed in the past, employing different techniques such as Neural Networks (see Refenes et al., 1997, Pantazopoulos et al., 1998, Zhang et al., 2002, and the special issue IEEE, 2001), Support Vector Machines (Cao and Tay, 2001 and 2003, Huang et al., 2005) or Hidden Markov Models (HMMs) (Hamilton and Susmel, 1994, Dueker, 1997, Shi and Weigend, 1997, Bengio et al., 2001, Liehr et al., 1999).

In more recent years, the research of financial econometricians has been mainly concentrated on the prediction of volatility in financial time series (Engle, 1982), rather than of the first moment of the returns or levels. In fact it has become widely accepted by the financial community that the returns (price variations) are unpredictable (see, for example, Samuelson, 1965, and Campbell et al., 1997). The idea is that, if a pattern exists, traders would take advantage of it, but it will be destroyed by the power of the market. Nevertheless this unappealing behavior could be somehow relaxed if we consider *sufficiently short* periods (e.g. a couple of weeks). Actually, it is plausible that common opinions about short periods (typically referred to as *local trends*) can influence the markets, determining the expected behavior in short periods. This is not in contrast with the previous statement: in fact it is true that the market would destroy the existing pattern, but it needs a brief period (a local trend) to detect it and to change its dynamics. This idea is the basis of the technical analysis (see, for example, Pring, 1991, and Edwards and Magee, 1992), where different heuristics-driven graphical techniques and simple indicators (climate, flow-of-funds, market indicators) were developed to detect and forecast short trends in financial markets. Moreover, a recent line of research of econometric literature (nowadays poorly explored) pays attention to the forecastability of the sign of the returns (market direction) (Christofferson and Diebold, 2006, Christofferson et al., 2006). The models used are very simple, being based on the probability that the standardized return is bigger or lower than a certain value and not on the direct modelization of the sign. Among the features of the direction-of-change in financial markets, two aspects are emphasized by Christoffersonn and Diebold (2006): 1) the nature of sign dependence is highly nonlinear and 2) the zero mean of the returns would render the sign unforecastable.

The approach presented in this paper could be collocated in the aforementioned context, and represents a probabilistic methodology (based on HMMs, which are nonlinear models) aimed at recognizing and forecasting the sign of financial local trends, in which the mean of the returns is positive or negative (dealing with sequences of *increases* or *decreases*). This aspect is of great practical importance, as it is related to investment decisions. Actually, what a broker expects from an analysis model is an aid to forecast if a certain share (or stock index) with a positive (negative) variation in the latest τ days, will continue to increase (decrease) in the next days. This problem, poorly considered by the researchers, needs specific models, able to capture the movement in the short time and the asymmetries between increase and decrease periods. In the proposed methodology this is accomplished by employing *two different* HMMs to explicitly, and separately, model the situations leading to increases or decreases. A novelty with respect to the mentioned (but recent) literature on the direction-of-change forecasting is that we modelize directly the sign of the returns, trying to capture the underlying nonlinear dependence structure. The models were trained with sequences that we trust be of increase (or decrease) — i.e. short sequences of T periods (few weeks of daily data), ending with τ increases (decreases).

When dealing with HMMs and financial series, a serious problem is to choose a proper distribution for the levels or the returns of the series; the difficulties arise because of the presence of high peaks, fat tails and higher kurtosis with respect to the Normal distribution; the typical solution is to increase the complexity of the distributions, assuming Student's t, GED, Gamma, etc. (see, for example, Taylor, 1999). In our approach we chose an opposite criterion: instead of adopting complicated distributions, we performed a preprocessing of the series, with the goal of removing unnecessary information – leading to a simplified representation. In particular, since we are interested in modelling the signs of the sequences, we transformed the original series in binary strings (encoding only increases or decreases), assuming then a simple binomial distribution associated with each state of the hidden Markov chain. In other words, the transformation of data in binary strings has a double advantage: it simplifies the hypotheses on the distribution of the time series and provides the direct modelization of the signs.

The proposed approach has been tested with encouraging results on the U.S.A. Dow Jones index and other shares of the same stock market with different risk. The forecasting and recognition capabilities of the proposed methodology have been assessed in different experimental conditions, and compared with models of Local Polynomial Terms (Kendall et al., 1983). The experiments assume more relevance noting that the training set (namely the sequences used to build the model) and the testing set are very large (around 2.5 years of daily data) and completely separated. In practice, the forecasts are performed for a very large horizon, which is generally a strong limitation of the usual time series models.

The remainder of the paper is organized as follow: in the next section the proposed methodology is detailed, whereas Section 3 contain the experimental evaluation performed on real data. In Section 4 we propose a solution to detect, with a probabilistic criterion, the sequences which can be classified as local trends, following our definition. Finally, in Section 5 conclusions are drawn and future perspectives are envisaged.

2 The proposed approach

In this section we describe the steps to transform the data and construct the double HMM model proposed in this paper. Before to describe the steps of the procedure, we will briefly recall the theory of HMMs, mainly for fix the notation employed in the rest of the paper. We refer to Rabiner (1989) for details on the HMM methodology.

2.1 Hidden Markov Models

A discrete-time hidden Markov model $\boldsymbol{\lambda}$ can be viewed as a Markov model whose states are not directly observed: instead, each state is characterized by a probability distribution function, modelling the observations corresponding to that state. More formally, a HMM is defined by the following entities (Rabiner, 1989): a set $S = \{S_1, S_2, \cdots, S_N\}$ of (hidden) states; a transition matrix $\mathbf{A} = \{a_{ij}\}$, where $a_{ij} \geq 0$ represents the probability of going from state S_i to state S_j; an emission matrix $\mathbf{B} = \{b(o|S_j)\}$, indicating the probability of emission of symbol o from state S_j; an initial state probability distribution $\boldsymbol{\pi} = \{\pi_i\}$, representing the probability of the first state $\pi_i = P[Q_1 = S_i]$.

For a sequence $\mathbf{O}$ and a HMM $\boldsymbol{\lambda}$, there is a standard recursive procedure able to compute the probability $P(\mathbf{O}|\boldsymbol{\lambda})$, and is called the *forward-backward* procedure (Rabiner, 1989). Given a set of observed sequences $\{\mathbf{O}_i\}$, the learning of the HMM parameters is usually performed using the well-known Baum-Welch algorithm (Rabiner, 1989), which is able to determine the parameters maximizing the likelihood $P(\{\mathbf{O}_i\}|\boldsymbol{\lambda})$.

2.2 The financial framework

Let us consider a time series representing the returns relative to a certain index, share, etc. The final purpose of the presented methodology is to recognize and forecast the sequences of the series that potentially end with a sub-sequence of τ signs of the same nature (all increases or all decreases). Let us call F_+ the set of sequences ending with τ positive signs, F_- the set of sequences ending with τ negative signs and $F_?$ the set of all the other sequences, which present an irregular alternation of the last τ signs. In practice the sequences belonging to F_+ can be considered as sequences of "reasonable sure increase", whereas the sequences belonging to F_- as "reasonable sure decrease"; we will call the sequences belonging to $F_?$ as "fluctuant".

The main feature of the proposed methodology is that the F_+ and F_- sets are modeled separately. Actually we have to observe that many studies in financial econometrics (see, for example, Hamilton and Susmel, 1994, Dueker, 1997) stress the fact that the volatility of financial time series is characterized by asymmetric behavior, because the quiet and turmoil periods have different length and persistence; in this framework a HMM approach seems suitable, each state representing a different regime of volatility. For example, in a 2-states HMM, the first state could represent the high volatility corresponding to the turmoil periods, whereas the second state could represent the low volatility corresponding to the quiet periods; the asymmetric transition probability matrix provides the different length and persistence of the two states along the full period studied. Similar hypotheses could be made for the sign of returns, emphasizing two degrees of asymmetry: first, the sequences belonging to F_+ and F_- have different behaviors, which can be pointed out estimating two different independent models; second, each set of sequences can show asymmetries also in each proper dynamics, represented by the alternation of the states with different length and persistence.

In the following subsections the proposed approach is presented. In particular, the three main parts of the methodology are fully detailed:

- the coding strategy; first, the series of returns is transformed in a binary string.
- the training procedure: to this aim the sequences (of short length T) belonging to F_+ and F_- are extracted from the full time series; then the two HMM models are trained (estimated) with F_+ and F_-, respectively. The number of states of each HMM model is automatically estimated from the training set.
- the testing method: how to employ the trained HMMs to correctly assign the testing sequences to F_+ or F_-.

2.3 Coding strategy

The goal of this phase is to transform the original series in a binary string. Discrete data have been used in several works dealing with financial time series; for example, in Frankel and Rose (1996) a logit model is used to analyze currency crises in the emerging markets, in Holthausen and Larcker (1992) to evaluate abnormal returns. As explained in the introduction, we performed this kind of preprocessing of the series for removing unnecessary information and for reaching a simplified representation. Since we are interested in modelling the signs of the sequences, transforming the original series in a binary string (where one symbol encodes the positive returns, the other the negative ones) seems the natural choice. It should be noted that using more than two symbols could be also reasonable: for example, five symbols could represent respectively the case of no significant variation, small increase, small decrease, large increase, large decrease. In this last case a proper quantization mechanism should be inserted. A preliminary experimentation, on the data described in Section 3, using *Vector Quantization* (VQ — Gersho and Gray, 1991) has been carried out, showing no significant variations.

2.4 Training

The training phase is aimed at building two models, one devoted to modelling the sequences belonging to F_+ and one for the sequences of F_-. The choice of τ is quite subjective, depending on the number of consecutive positive days considered as an interesting gain by the investor, and could have a not negligible impact on the performances. In the experimental results proposed in Sect. 3 we will evaluate different configurations.

The two sets of sequences F_+ and F_- are used to train two discrete Hidden Markov Models (i.e. HMMs where the emission probability $b(o|S_j)$ is a binomial distribution), called $\boldsymbol{\lambda}_+$ and $\boldsymbol{\lambda}_-$ respectively. The training is performed using the standard Baum Welch re-estimation method (Rabiner, 1989), which is stopped at likelihood convergence. As described in greater detail in the following, each HMM is carefully initialized, and the number of states is automatically estimated. Actually, the initialization process heavily affects the resulting model estimate, as the likelihood function is highly multi-modal and the training procedure converges to the nearest maxima. In this paper this problem has been faced with an effective, even if quite standard approach (see for example McLachlan and Peel, 2000): the training has been repeated several times starting from different random conditions, choosing the model leading to the highest likelihood.

The second practical but fundamental issue to be addressed when using an HMM is the determination of the number of states: here we used the *Akaike Information Criterion* (AIC — Akaike, 1974). This choice is reasonable, since it has been shown in Psaradakis and Spagnolo (2003) that, in this context, this criterion is more suitable at determining the optimal number of states[1] with respect to other criteria, like BIC. Given $\mathcal{O}$ the observed data-set, under the AIC criterion the optimal number of states is the maximizer of AIC(k):

$$\widehat{k}_{\text{AIC}} = \arg\max_k \text{AIC}(k) \qquad \text{with AIC}(k) = 2\log p(\mathcal{O}|\widehat{\boldsymbol{\lambda}}_k) - 2N_k$$

where $\widehat{\boldsymbol{\lambda}}_k$ denotes the Maximum Likelihood estimate of the model with k states, and N_k is the total number of free parameters of $\widehat{\boldsymbol{\lambda}}_k$.

2.5 Testing

Given an unclassified sequence O, the testing phase is very straightforward, and is obtained by feeding the sequence to both the models, computing the probability $P(O|\boldsymbol{\lambda}_+)$ and $P(O|\boldsymbol{\lambda}_-)$. The first represents the probability that the sequence has been generated by the $\boldsymbol{\lambda}_+$ or, equivalently, that the sequence belongs to F_+, whereas the second is the probability that the sequence has been generated from $\boldsymbol{\lambda}_-$, or that it belongs to F_-. The decision is subsequently performed by using the *Bayesian decision rule* (Duda et al., 2001), which claims to assign an unknown sequence to the class whose posterior is maximized. Any monotonic transformation of $P(O|\boldsymbol{\lambda})$ does not affect the classification results. For computational reasons (see Rabiner, 1989) we compute the logarithm of this probability. Assuming a priori equiprobable classes, the rule reduces to the *Maximum Likelihood criterion*, that is:[2]

$$Class(O) = \begin{cases} F_+ & \text{if } \log P(O|\boldsymbol{\lambda}_+) > \log P(O|\boldsymbol{\lambda}_-) \\ F_- & \text{otherwise} \end{cases} \tag{1}$$

3 Experimental Results

The proposed approach has been tested using the Dow Jones index and other related shares: we employed daily close prices from 30 November 1995 to 5 February 2001, adjusted for dividends and splits — Yahoo Finance source. The models were trained using the period from December 1995 to June 1998, whereas the remaining part of the series has been used for testing.

The experimental evaluation has been carried out pursuing four different goals: in the first part the analysis was restricted to the Dow Jones index, with the goal of testing recognition and forecasting capabilities of the proposed approach with a large range of parameters' values. The obtained results have been then validated in the second part by using other shares, employing the best parameters' configuration found in the first analysis. The third part was devoted to compare the proposed methodology with another technique.

[1] Actually, the proof, involving Monte Carlo simulation, was made for Autoregressive Markov Switching Models, that are dynamic models with parameters subject to Markov regime switching.

[2] We call $Class(O)$ the classification of O resulting from the decision rule.

Table 1: Number of training sequences and best number of states (AIC) for different values of parameters.

Sequence length T	Repeated symbols τ	Number of train seq. in F_+	Number of train seq. in F_-	States of model $\boldsymbol{\lambda}_+$	States of model $\boldsymbol{\lambda}_-$
10	3	107	71	8	8
10	4	53	26	5	4
10	5	28	8	3	2
15	3	107	70	6	4
15	4	53	26	5	3
15	5	28	8	4	3
20	3	107	70	4	4
20	4	53	26	5	3
20	5	28	8	4	3
25	3	107	70	2	4
25	4	53	26	2	3
25	5	28	8	5	3

Finally, in the fourth part a discussion on how to detect sequences belonging to $F_?$ has been proposed (it is described in the next section).

3.1 Experiments on Dow Jones

A first large scope analysis has been carried out with the Dow Jones index, testing different values of the parameters – 4 different sequences lengths (T=10, 15, 20, 25) and 3 different τ (τ=3, 4, 5). Clearly these two parameters influence the number of sequences used to build the models. This could be noted by looking at table 1, where these information are displayed together with the optimal number of states determined by the AIC criterion for all models. It is evident that the number of increases is higher than the number of decreases, this suggesting an asymmetry between the two typologies of series. This is confirmed by considering that the corresponding optimal models show always a different number of states (except with T=10 and τ=3). It is also interesting to note that small models are preferred, once again confirming the well known Occam razor principle (Thorburn, 1915).

In order to test both the recognition and the forecasting capabilities of the method, two different experiments have been carried out.

3.1.1 Recognition

The recognition capability of the proposed methodology was tested by classifying all the sequences of reliable increase and reliable decrease extracted from the testing set (disjoint from the training). Each sequence has been fed to the two trained HMMs, assigning it to the class whose model shows the highest log likelihood

Table 2: Accuracies for experiment 1 and 2 for different parameters configurations.

Sequence length T	Repeated symbols τ	Number of test seq. in F_+	Number of test seq. in F_-	Recognition Accuracy (Exp 1)	Forecast -1 (Exp 2)	Forecast -2 (Exp 2)	Forecast -3 (Exp 2)
10	3	78	80	100.00%	100.00%	100.00%	56.96%
10	4	39	29	100.00%	100.00%	100.00%	100.00%
10	5	18	8	100.00%	100.00%	100.00%	100.00%
15	3	78	78	96.15%	87.18%	71.15%	42.31%
15	4	39	28	94.03%	89.55%	80.60%	71.64%
15	5	18	8	100.00%	96.15%	96.15%	80.77%
20	3	78	78	83.97%	72.44%	57.69%	43.59%
20	4	39	28	89.55%	82.09%	73.13%	61.19%
20	5	18	8	96.15%	96.15%	88.46%	65.38%
25	3	78	77	57.42%	53.55%	46.45%	42.58%
25	4	39	28	76.12%	67.16%	61.19%	52.24%
25	5	18	8	88.46%	80.77%	73.08%	69.23%

(rule (1)).

The percentages of correct classification (accuracy) have been computed for different parameters configurations: results are reported in the fifth column of table 2. Results are really satisfactory for large part of the configurations: in most cases the percentage of correct classification is high (more than 80% if we exclude the case $T = 25$). Note that in general the performances worsen when increasing the length of the sequence, especially with $\tau = 3$. This confirms the aforesaid intuition: longer sequences mislead the system, since the market is able to detect and correspondingly destroy the pattern.

3.1.2 Forecasting

The goal of this experiment was to assess the forecasting performances of the proposed approach, namely the capability of predicting the increase or decrease before its happening. To this aim, in this experiment, the sequences of reliable increase or decrease extracted from the testing set have been truncated and fed to the models to be classified; the classification result represents an attempt to predict in advance the behavior of the whole sequence (i.e. by just observing a part of it).

Results are displayed in table 2 (last three columns), for different temporal horizons. More precisely, "Forecast -k" means that the forecasting is performed with a horizon of k days: for example, for sequence length equal to 15, "Forecast -2" means that only the first 13 symbols of each sequence are presented to the system, which should forecast the sign of the whole sequence (15 symbols long) without knowing the 14th and 15th values. From the table one can notice that results are very satisfactory for short sequences: it seems that two or three weeks of daily closures are sufficient to properly characterize the increases and the decreases. The considerations made for the previous experiment are valid again; the performance is particularly good

Table 3: Number of training sequences and best number of states for different indices.

Index	Number of train seq. in F_+	Number of train seq. in F_-	States of increase model	States of decrease model
GE	16	20	4	4
IBM	17	17	3	2
BGP	7	25	2	4
CGI	12	38	3	4

Table 4: Accuracies for experiment 1 and 2 for different indices.

Index	Number of test seq. in F_+	Number of test seq. in F_-	Recognition Accuracy (Exp 1)	Forecast -1 (Exp 2)	Forecast -2 (Exp 2)	Forecast -3 (Exp 2)
GE	16	14	100.00%	100.00%	100.00%	83.33%
IBM	19	20	100.00%	100.00%	100.00%	100.00%
BGP	6	33	100.00%	100.00%	100.00%	100.00%
CGI	2	38	100.00%	100.00%	100.00%	100.00%

for short sequences with τ equal to 4 or 5, worsening when enlarging the horizon of forecasts.

3.2 Experiments on other share indices

In order to get some more insights into the methodology, the previous experiments have been repeated using some other shares with different risks, belonging to the components of the Dow Jones index: GE (General Electric) and IBM (International Machines Corp.) (low risk shares), BGP (Borders Group) and CGI (Commerc Group) (high risk shares). The best parameter configuration derived from the Dow Jones analysis has been used, i.e. $T = 10$ and $\tau = 5$. Experiments details and results for recognition and forecasting experiments are reported in table 3 and 4, respectively. Observing the table 3 it is evident that also in this case the models $\boldsymbol{\lambda}_+$ and $\boldsymbol{\lambda}_-$ have different number of states, confirming the asymmetry of the phenomenon. From table 4, it is interesting to notice that the shares show an optimal performance in the recognition and forecast of the reliable sequences, and the different risk does not affect the forecast accuracy.

3.3 Comparative analysis

The proposed approach has been compared with a local polynomial trend (LP model – Kendall et al., 1983):

$$y_t = \alpha + \sum_{i=1}^{p_h} \beta_i t^i \tag{2}$$

where y_t represents the observation (the level of the series) at time t and h refers to the model of increase (h=1) or decrease (h=2).

Table 5: Experiments 1 and 2 with local polynomial trends.

Share	Recognition Accuracy (Exp 1)	Forecast -1 (Exp 2)	Forecast -2 (Exp 2)	Forecast -3 (Exp 2)
GE	76.67%	86.67%	76.67%	56.67%
IBM	92.31%	89.74%	82.05%	64.10%
BGP	46.15%	79.49%	61.54%	43.59%
CGI	63.41%	73.17%	56.10%	51.22%

This choice seems to be adequate for this kind of analysis: actually, the technical analysts do not employ statistical models but empirical evidence for their operations – see Pring (1991). Different models have been estimated using different degrees p_1 and p_2 (ranging from 1 to T - 4), finally choosing the configuration maximizing the prediction accuracy.

In this case, the coefficients α and β_i were estimated (via ordinary least squares) for each local trend to be recognized (or predicted). In practice, we have not utilized a training set, but the local trends of interest (eliminating the latest k observations in the forecasting experiments). As a matter, the information used in this analysis is more accurate with respect the HMM experiments. Then we used the estimated coefficients to estimate the observations of the local trend and verified if the estimated signs of the returns are equal to the true ones.

The analogous experiments described in the previous section have been repeated with these models; the results are shown in Table 5. By comparing the results with Table 4, it could be noted the great increase in accuracy gathered by the HMM modelling strategy proposed in this paper. In particular, the 3-steps forecasting performance of the local polynomial trends is almost completely random.

4 Detection of fluctuant sequences

In practical cases, given a generic sequence, the goal of an investor is to get some suggestions for the solution of the following problems:

1. Does the sequence present a clear and defined trend? (Namely is it of reliable increase or of reliable decrease?)

2. If so, does the sequence represent a reliable increase or a reliable decrease? Moreover, can we detect such situation in advance?

The main purpose of this paper was to answer to the second question. The experiments illustrated in the previous section showed that once identified such sequences our approach is really appropriate in both recognition and forecasting. Nevertheless, explicitly addressing the first issue is crucial for practical applications. In this section we provide a possible solution to this aspect by introducing in the proposed methodology a reject rule. The key idea is that it is possible to reject the decision if the system is not confident enough about a classified sequence; the rejection of a decision would then correspond to the assignation of the sequence to the set $F_?$, namely to the set of "fluctuant" sequences. In the Pattern Recognition area, the most widely employed rejection rule is the Chow rule (the basic version could be found in Chow, 1957, some more complex in Fumera et al., 2000), which indicates to reject a decision if the maximum posterior probability is under a given threshold. This rule is optimal in a Bayesian sense, leading to the optimal reject-error tradeoff, but assumes the complete knowledge of the posterior. Moreover it is defined in a multi-class context, and could not completely exploit the dichotomy of the problem. For these reasons a different rule is introduced here, following a different rationale: instead of thresholding the posterior, we propose to threshold the difference between the log likelihoods of the two models. In details, we introduce the concept of *confidence of a classification*, defined as

$$Confidence(O) = |\log P(O|\boldsymbol{\lambda}_+) - \log P(O|\boldsymbol{\lambda}_-)| \quad (3)$$

If the system does not show a clear evidence in favor of a certain class, i.e. $Confidence(O) < \theta$ (for a chosen threshold θ), the sequence is not assigned neither to F_+ nor to F_-, but it is rejected (namely it is assigned to $F_?$).

The choice of the threshold θ is obviously crucial, and it is linked to the level of confidence desired by the user. Here we propose a simple and automatic way of setting this parameter on the basis of the training set. In particular we divide the training set in two subsets, one containing the sequence with a clear trend (i.e. both the sequences of reliable increase and reliable decrease), and one with the others. Then the confidence (3) is computed for both the sets and the two empirical pdf's are estimated: one represents the probability of the confidence given that the sequence has a clear trend, while the other is the probability of the confidence given that the sequence is a fluctuating sequence. The threshold is then computed in a Bayesian way, by considering the point where the two pdfs intersect: this represents the minimum error point. The idea is briefly sketched in Fig. 1: the continuous line represents the empirical distribution of the statistic (3) for the sequences of length $T = 10$ of the training set belonging to $F_?$ and the dot line the analogous distribution for the sequences with $T = 10$ and $\tau = 5$ of the training set belonging to F_+ and F_-. The intersection point of the two curves detects the threshold θ (equal to 3.1868). All the sequences of the testing set with $Confidence(O) < \theta$ will be considered as belonging to $F_?$, whereas the others will be classified in F_+ and F_- by the decision rule (1).

In Table 6 we show the percentage of the right rejections in an experiment analogous to Experiment 1, but considering all the possible sequences. These accuracies represent the number of times the system is able to correctly rejecting a sequence belonging to $F_?$. The percentage is generally high, in particular for the returns relative to the four shares. Again we have to stress that results are obtained on the testing set, whereas models and the employed threshold have been determined on the (disjoint) training set.

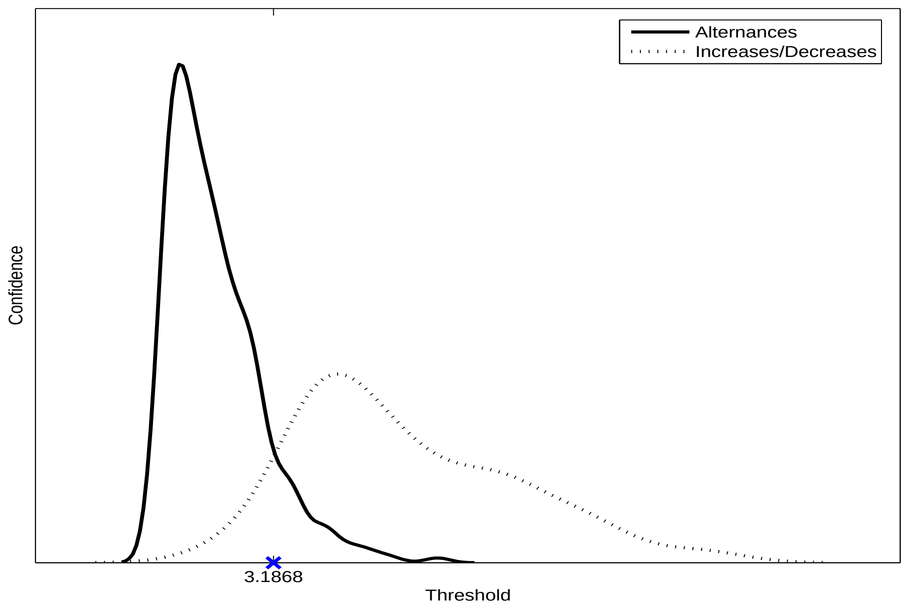

Figure 1: Empirical distributions of the confidence statistics for $T = 10$ and $\tau = 5$ and detection of the threshold.

Table 6: Percentage ratios of rejected sequences over the number of sequences belonging to $F_?$ in the testing set.

Sequence length T	Repeated symbols τ	Correctly Rejected sequences
10	3	76.04%
10	4	84.18%
10	5	93.07%
15	3	66.93%
15	4	67.80%
15	5	87.93%
20	3	79.32%
20	4	74.02%
20	5	85.12%
25	3	97.77%
25	4	80.86%
25	5	83.06%
GE		95.09%
IBM		85.85%
BGP		91.29%
CGI		93.31%

5 Final remarks

In this paper a novel approach for recognizing and forecasting brief sequences of series relative to financial markets has been proposed. The model explicitly and directly exploit the natural asymmetry present in the market by training two separate models, one for the increase situation and one for the decrease. The resulting methodology is quite simple, being based on the automatic selection of the number of the Markovian states (using the AIC) and being fully data driven, not using explicative variables.

Future efforts will be devoted to extend the presented model. An immediate attempt would be the enrichment of the methodology with the linking of the some external information; e.g., it is plausible that the movements in a leading market (such as the U.S. market) influence the movements in the other markets. There are many possibilities: inserting in the transition probabilities of the HMM a dependence on the value of the other markets, as in Diebold et al. (1994), or modelling multiple sequences (namely different indices or shares) with the same Markov Chain or a multiple Markov Chain, with a mutual influence among the states of different variables (as in Otranto, 2005), or finally by modelling interactions between couple of processes (as in the Coupled Hidden Markov Models Brand et al., 1997).

Acknowledgements

The authors thank very much Giangiacomo Milia for preliminary experiments on the series and Giampiero M. Gallo for the useful comments.

References

[1] Y. Bengio, V.-P. Lauzon, and R. Ducharme (2001) Experiments on the application of IOHMMs to model financial returns series. *IEEE Transaction on Neural Networks*, 12(1):113–123.

[2] M. Brand, N. Oliver, and S. Pentland (1997). Coupled hidden markov models for complex action recognition. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition.*

[3] J.Y. Campbell, A.W. Lo, and A.C. MacKinlay (1997). *The Econometrics of Financial Markets.* Princeton University Press: Princeton.

[4] L. Cao and F. Tay (2001). Financial forecasting using support vector machines. *Neural Computing and Applications*, 10:184–192.

[5] L.J. Cao and F.E.H. Tay (2003). Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Trans. on Neural Networks*, 14(6):1506–1518.

[6] C.K. Chow. (1957) An optimum character recognition system using decision functions. *IRE Trans. on Electronic Computers*, 6:247–254.

[7] P.F. Christofferson and F.X. Diebold (2006). Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science*, 52(8):1273–1287.

[8] P.F. Christofferson, F.X. Diebold, R.S. Mariano, A.S. Tay, and Y.K. Tse (2006). Direction-of-change forecasts based on conditional variance, skewness and kurtosis dynamics: international evidence. *PIER Working Paper*, pages 06–016.

[9] F.X. Diebold, J.H. Lee, and G.C. Weinbach (1994). Regime switching with time-varying transition probabilities. In P. Hargreaves, editor, *Nonstationary Time Series Analysis and Cointegration*, pages 283–302. Oxford University Press.

[10] R.O. Duda, P.E. Hart, and D.G. Stork (2001). *Pattern Classification.* John Wiley and Sons, second edition.

[11] J.M. Dueker (1997). Markov switching in garch processes and mean-reverting stock market volatility. *Journal of Business and Economic Statistics*, 15:26–34.

[12] R.D. Edwards and J. Magee (1992). *Technical Analysis of Stock Trends.* J. Magee Institute: Boston.

[13] R.F. Engle (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of u.k. inflation. *Econometrica*, 50:987–1008.

[14] J.A. Frankel and A.K. Rose (1996). Currency crashes in emerging markets: an empirical treatment. *Journal of International Economics*, 41:351–366.

[15] G. Fumera, F. Roli, and G. Giacinto (2000). Reject option with multiple thresholds. *Pattern Recognition*, 33:2099–2101.

[16] A. Gersho and R. Gray (1991). *Vector Quantization and Signal Compression.* Kluwer Acad. Press.

[17] J.D. Hamilton and R. Susmel (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, 64:307–333.

[18] R.W. Holthausen and D.F. Larcker (1992). The prediction of stock returns using financial statement information. *Journal of Accounting and Economics*, 15:373–412.

[19] W. Huang, Y. Nakamori, and S.-Y. Wang (2005). Forecasting stock market movement direction with support vector machine. *Computers and Operations Research*, 32(10):2513–2522.

[20] IEEE (2001). Special issue on neural networks in financial engineering. *IEEE Transactions on Neural Networks*, 12(4).

[21] G. Kendall, A. Stuart, and J.K. Ord (1983). *The advanced Theory of Statistics.* Griffin: London.

[22] S. Liehr, K. Pawelzik, J. Kohlmorgen, S. Lemm, and K.-R. Müller (1999). Hidden markov mixtures of experts for prediction of non-stationary dynamics. In *IEEE Proc. of Neural Networks for Signal Processing IX*, pages 195–204.

[23] G. McLachlan and D. Peel (2000). *Finite mixture models*. John Wiley and Sons.

[24] E. Otranto (2005). The multi-chain markov switching model. *Journal of Forecasting*, 24(7):523–537.

[25] K.N. Pantazopoulos, L.H. Tsoukalas, N.G. Bourbakis, M.J. Brun, and E.N. Houstis (1998). Financial prediction and trading strategies using neurofuzzy approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 28(4):520–531.

[26] M.J. Pring (1991). *Technical Analysis Explained*. McGraw-Hill, 3 edition.

[27] Z. Psaradakis and Spagnolo (2003). On the determination of the number of regimes in Markov-Switching autoregressive models. *Journal of Time Series Analysis*, 24:237–252.

[28] L. Rabiner (1989). A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proc. of IEEE*, 77(2):257–286.

[29] A.P.N. Refenes, A.N. Burgess, and Y. Bentz (1997). Neural networks in financial engineering - a study in methodology. *IEEE Transactions on Neural Networks*, 8(6):1222–1267.

[30] P.A. Samuelson (1965). Proof that properly anticipated prices fluctuate randomly. *Ind. Manage. Rev.*, 6:41–49.

[31] S. Shi and A.S. Weigend (1997). Taking time seriously: Hidden markov experts applied to financial engineering. In *Proc. of IEEE/IAFE 1997 Conf. on Computational Intelligence for Financial Engineering*, pages 244–252.

[32] S.J. Taylor (1999). Markov processes and the distribution of volatility: a comparison of discrete and continuous specifications. *Philosophical Transactions: Mathematical, Phisical and Engineering Sciences*, 357:2059–2070.

[33] W. M. Thorburn (1915). Occam's razor. *Mind*, 24:287–288.

[34] M. Zhang, S. Xu, and J. Fulcher (2002). Neuron-adaptive higher order neural-network models for automated financial data modeling. *IEEE Transactions on Neural Networks*, 13(1):188–204.

Ultimi Contributi di Ricerca CRENoS

I Paper sono disponibili in: http://www.crenos.it

08/02 *Juan de Dios Tena, Edoardo Otranto,* "A Realistic Model for Official Interest Rates Movements and their Consequences"

08/01 *Edoardo Otranto,* "Clustering Heteroskedastic Time Series by Model-Based Procedures"

07/16 *Sergio Lodde,* "Specialization and Concentration of the Manufacturing Industry in the Italian Local Labor Systems"

07/15 *Giovanni Sulis,* "Gender Wage Differentials in Italy: a Structural Estimation Approach"

07/14 *Fabrizio Adriani, Luca G. Deidda, Silvia Sonderegger,* "Over-Signaling Vs Underpricing: the Role of Financial Intermediaries In Initial Public Offerings"

07/13 *Giovanni Sulis,* "What Can Monopsony Explain of the Gender Wage Differential In Italy?"

07/12 *Gerardo Marletto,* "Crossing the Alps: Three Transport Policy Options"

07/11 *Sergio Lodde* "Human Capital And Productivity Growth in the Italian Regional Economies: a Sectoral Analysis"

07/10 *Axel Gautier, Dimitri Paolini,* "Delegation, Externalities and Organizational Design"

07/09 *Rinaldo Brau, Antonello E. Scorcu, Laura Vici,* "Assessing visitor satisfaction with tourism rejuvenation policies: the case of Rimini, Italy"

07/08 *Dimitri Paolini,* "Search and the firm's choice of the optimal labor contract"

07/07 *Giacomo Carboni,* "Shape of U.S. business cycle and long-run effects of recessions"

07/06 *Gregory Colcos, Massimo Del Gatto, Giordano Mion and Gianmarco I.P. Ottaviano,* "Productivity and firm selection: intra-vs international trade"

07/05 *Silvia Balia,* "Reporting expected longevity and smoking: evidence from the share"

07/04 *Raffaele Paci, Stefano Usai,* "Knowledge flows across European regions"

07/03 *Massimo Del Gatto, Giordano Mion and Gianmarco I.P. Ottaviano,* "Trade Integration, firm selection and the costs of non-europe"

07/02 *Vittorio Pelligra,* "Intentions, trust and frames: a note on sociality and the theory of games"

07/01 *Oliviero A. Carboni and Giuseppe Medda,* "Government size and the composition of public spending in a neoclassical growth model"

06/17 *Susana Iranzo, Fabiano Schivardi and Elisa Tosetti,* "Skill dispersion and firm productivity: an analysis with employer-employee matched data"

06/16 *Luigi Guiso and Fabiano Schivardi,* "What determines entrepreneurial clusters?"

06/15 *Vittorio Pelligra,* "Trust Responsiveness: on the dynamics of fiduciary interactions"

06/14 *Rinaldo Brau, Alessandro Lanza and Francesco Pigliaru,* "How fast are small tourist countries growing? The 1980-2003 evidence"

06/13 *Emanuela Marrocu and Raffaele Paci,* "The effects of public capital on the productivity of the italian regions"

www.crenos.it