Analyzing the sign of financial local trends via Hidden Markov Models

Manuele Bicego, Enrico Grosso and Edoardo Otranto

The problem of forecasting financial time series has received great attention in the past. In this paper we deal with 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. The approach has been tested by different experiments with the Dow Jones index.

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. In more recent years, the research of financial econometricians has been mainly concentrated on the prediction of volatility in financial time series, 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. 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), 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.

The approach presented in this paper could be collocated in the aforementioned context, and represents a probabilistic methodology (based on Hidden Markov Models - HMM hereafter) 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 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).

For this purpose we transform 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 remainder of the paper is organized as follow: in the next section we briefly describe the proposed methodology, 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.

2 The Proposed Approach

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, 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.

The approach consists of three parts (details are in Bicego et al., 2008a and 2008b):

- *coding:* the series of returns is transformed in a binary string (0 for negative variations, 1 for positive variation).
- *training:* 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 estimated (trained) with F_+ and F_-, respectively. The number of states of each HMM model is automatically detected from the training set, using the AIC criterion.
- *testing:* employ the trained HMMs to correctly assign the testing sequences to F_+ or F_-. The assignment is made using the maximum likelihood criterion.

3 Experimental Results

The proposed approach has been tested using the Dow Jones index: 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

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 λ_+	States of model λ_-
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

been used for testing.

A first large scope analysis has been carried out 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 Recognition (Experiment 1)

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. 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 τ = 3. This confirms the aforesaid intuition: longer sequences mislead the system, since the market is able to detect and correspondingly destroy the pattern.

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%

3.2 Forecasting (Experiment 2)

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

4 Detection or fluctuant sequences

The experiments illustrated in the previous section showed that, once identified sequences of reliable increase or decrease, our approach is really appropriate in both recognition and forecasting. What happens if the sequence can not be classified as reliable increase or decrease? 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. 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* (CO), defined as the difference between the log-likelihood of the HMM for F_+ and the log-likelihood of the HMM for F_-. If the system does not show a clear evidence in favor of a certain class, i.e. $CO < \theta$ 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 CO 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 CO 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

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

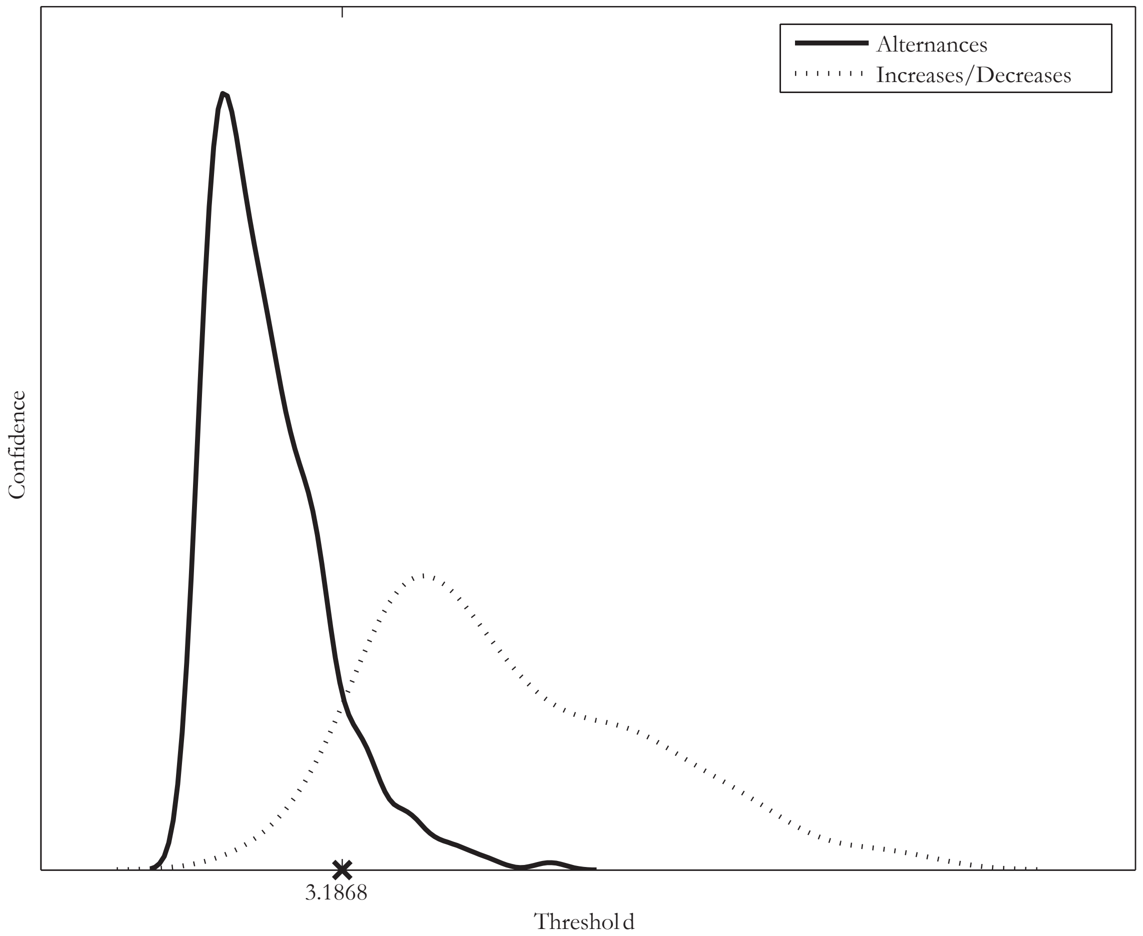

Table 3. Percentage ratios of rejected sequences over the number of sequences belonging to $F_?$ in the testing set.

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%

θ (equal to 3.1868). All the sequences of the testing set with $CO < \theta$ will be considered as belonging to $F_?$, whereas the others will be classified in F_+ and F_- by the decision rule used in the training and testing steps.

In Table 3 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. 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.

About the Authors

Manuele Bicego is Assistant Professor in Computer Science at University of Verona (home page http://profs.sci.univr.it/ ~ bicego/);

Enrico Grosso is Full professor in Computer Science at University of Sassari;

Edoardo Otranto is Associate Professor in Statistics at University of Sassari (home page http://economia.uniss.it/docenti/ otranto edoardo/ home en.htm).

References

[1] M. Bicego, E. Grosso, and E. Otranto (2008a). Recognizing and Forecasting the Sign of Financial Local Trends using Hidden Markov Models, CRENoS Working Paper 2008/03.

[2] M. Bicego, E. Grosso, and E. Otranto (2008b). A Hidden Markov Model approach to classify and predict the sign of financial local trends Proc. IAPR Int. work-

shop on Statistical Techniques in Pattern Recognition (SPR2008), Lecture Notes in Computer Science 5342, 852-861.
[3] P.F. Christofferson and F.X. Diebold (2006). Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science* 52, 1273{1287.
[4] P.F. Christofferson, F.X. Diebold, R.S. Mariano, A.S. Tay, and Y.K. Tse (2007). Direction-of-change forecasts based on conditional variance, skewness and kurtosis dynamics: international evidence. *The Journal of Financial Forecasting* 1, 3−24 .
[5] J.M. Dueker (1997). Markov switching in GARCH processes and mean-reverting stock market volatility. *Journal of Business and Economic Statistics* 15, 26−34.
[6] M.J. Pring (1991). *Technical Analysis Explained*. McGraw-Hill, 3-th edition.
[7] W. M. Thorburn (1915). Occam's razor. *Mind* 24, 287−288.