

# A Hidden Markov Model Approach to Classify and Predict the Sign of Financial Local Trends

Manuele Bicego, Enrico Grosso, and Edoardo Otranto

DEIR - University of Sassari  
via Torre Tonda, 34 - 07100 Sassari (Italy)  
{bicego,grosso,eotranto}@uniss.it

**Abstract.** In the field of financial time series analysis it is widely accepted that the returns (price variations) are unpredictable in the long period [1]; nevertheless, this unappealing constraint could be somehow relaxed if *sufficiently short* time intervals are considered. In this paper this alternative scenario is investigated with a novel methodology, aimed at analyzing short (local) financial trends for predicting their sign (increase or decrease). This peculiar problem needs specific models – different from standard techniques used for estimating the volatility or the returns – able to capture the asymmetries between increase and decrease periods in the short time. This is achieved by modeling directly the signs of the local trends using *two* separate Hidden Markov models, one for positive and one for negative trends. The approach has been tested with different financial indexes, with encouraging results also in comparison with standard methods.

## 1 Introduction

The analysis of financial signals represents a widely investigated research area, where many different methodologies have been proposed in the past, employing different techniques such as Neural Networks (see for example the special issue [2]), Support Vector Machines [3] or Hidden Markov Models (HMMs) [4,5,6].

Recently, the efforts of financial econometricians have been mainly concentrated on the prediction of volatility in financial time series [7], rather than on the price variations. In fact it has become widely accepted by that community that the price variations are unpredictable [1], following the idea that if a pattern exists it should be destroyed by the power of the market. Nevertheless this unappealing constraint could be somehow relaxed if *sufficiently short* periods are considered, as a couple of weeks: prior to the deletion of a pattern, the market 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, [8]), 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.

This paper addresses the aforementioned problem, proposing a methodology based on HMMs [9], aimed at recognizing and forecasting the sign of financial

local trends. The direct forecasting of the sign – with respect to the real value – represents a quite novel trend in the econometric literature (for example in [10] a set of heuristics and trivial models are used to this aim). The main feature of the proposed approach is represented by the explicit encoding of the asymmetry present in the problem, obtained by employing *two different* HMMs to separately model the situations leading to increases or decreases. This is substantially different from standard financial approaches based on HMMs, where just one model is employed [4,11]; actually such single model is able to capture only the implicit level of asymmetry present in the signal (represented with the hidden states). On the contrary, our approach considers and models the separate signals using *two models*, highly and explicitly emphasizing the asymmetries typically present in this kind of data. Another key feature is that, instead of hypothesizing a distribution on the levels or the returns of the series (which is actually a crucial problem), we model directly the signs, transforming the data in discrete binary sequences (modeled by discrete HMMs) representing positive variations (increases) and negative variation (decreases) with respect to the previous period.

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  $\tau$  increases (decreases). Each HMM training has been carefully initialized, while the number of states has been directly estimated from data using the Akaike Information Criterion (AIC – [12]). Recognition and prediction of signs are made on the basis of the likelihoods of the models; an automatic threshold-based reject rule has been also introduced, able to characterize “uncertainty” zones (zones neither of decrease nor of increase).

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; a comparison with alternative schemes has been also carried out.

## 2 The Proposed Approach

In this section the proposed approach is presented. Let us call  $F_+$  the set of sequences ending with  $\tau$  positive signs (sequences of “reasonable sure increase”),  $F_-$  the set of sequences ending with  $\tau$  negative signs (sequences of “reasonable sure decrease”) and  $F_?$  the set of all the other sequences, which present an irregular alternation of the last  $\tau$  signs. The choice of  $\tau$  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.

The main feature of the proposed methodology is that the  $F_+$  and  $F_-$  sets are modeled separately. Actually many studies in financial econometrics (see, for example, [11,4]) 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 *single* HMM seems suitable, each state representing a different regime of volatility. The asymmetric transition probability matrix provides the different length and persistence of the

states along the full period studied. A similar reasoning could be applied when modeling the signs. Even more, the asymmetry could be furthermore stressed by explicitly emphasizing two degrees of asymmetry: first, the sequences belonging to  $F_+$  and  $F_-$  have different behaviors (exploited by estimating two different independent models); second, each set of sequences can show asymmetries also in each proper dynamics (exploited by different states in each model).

## 2.1 HMM Definition

In this section, HMMs are briefly introduced, mainly to set up notation. A discrete-time hidden Markov model  $\lambda$  is defined by the following entities [9]:

- a set  $S = \{S_1, S_2, \dots, S_N\}$  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  $\pi = \{\pi_i\}$ , representing the probability of the first state  $\pi_i = P[Q_1 = S_i]$

For a sequence  $\mathbf{O}$  and an HMM  $\lambda$ , there is a standard recursive procedure, called the *forward-backward* procedure [9], able to compute the probability  $P(\mathbf{O}|\lambda)$ . 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 [9], which is able to determine the parameters maximizing the loglikelihood  $\log P(\{\mathbf{O}_i\}|\lambda)$ .

## 2.2 Coding

When dealing with HMMs and financial series, a serious problem is to choose a proper distribution (namely a proper emission function) for the levels or the returns of the series, difficult task for 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. In our approach we overcame this critical choice by modeling directly the signs of the sequences, transforming the original series in binary strings (series of increase or decrease). Discrete data have been used in several works dealing with financial time series [13,14]. It should be noted that using more than two symbols could be also reasonable: in this last case a proper quantization mechanism should be inserted<sup>1</sup>.

## 2.3 Training

The training phase is aimed at building two models, one devoted to modeling the sequences belonging to  $F_+$  and one for the sequences of  $F_-$ . The two sets of

<sup>1</sup> A preliminary experimentation using *Vector Quantization* has been carried out, showing no significant variations.

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  $\lambda_+$  and  $\lambda_-$  respectively. The training is performed using the standard Baum Welch re-estimation method [9], which is stopped at likelihood convergence. In order to deal with the initialization issue, the training has been repeated several times starting from different random conditions, choosing the model leading to the highest likelihood. Following findings in a related context [15], the number of states has been chosen with the *Akaike Information Criterion* (AIC - [12])<sup>2</sup>

### 3 Experimental Results

The proposed approach has been tested using the Dow Jones index (DJ) and 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). 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.

A first large scope analysis (not shown here) was carried out with the Dow Jones index, in order to find the most suitable values of the parameters. The test was carried with 4 different sequence lengths ( $T=10, 15, 20, 25$  - it has poor financial sense to perform the analysis outside the range two - five weeks) and 3 different  $\tau$  ( $\tau=3, 4, 5$  - we stopped the analysis at  $\tau = 5$  since no sequences with  $\tau = 6$  were present in the dataset.): best results were obtained with  $T = 10$  and  $\tau = 5$ . We noticed 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. In table 1 the number of sequences used to build the models 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 often really different than the number of decreases, this suggesting an asymmetry between the two typologies of signals. This is confirmed by considering that the corresponding optimal models show almost always a different number of states.

In order to test both the recognition and the forecasting capabilities of the method, two different experiments have been carried out. 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 (Bayes rule). The number of correct classifications are reported in the second column of table 2.

<sup>2</sup> 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.

**Table 1.** Number of training sequences and best number of states for different indices

Index	Train seq		States	
	$F_+$	$F_-$	$\lambda_+$	$\lambda_-$
DJ	28	8	3	2
GE	16	20	4	4
IBM	17	17	3	2
BGP	7	25	2	4
CGI	12	38	3	4

**Table 2.** Recognition and forecasting accuracies of the proposed approach for different indices (correct/total)

Index	Recogn.	Forecast	Forecast	Forecast
	Accuracy	-1	-2	-3
DJ	26/26	26/26	26/26	26/26
GE	30/30	30/30	30/30	25/30
IBM	39/39	39/39	39/39	39/39
BGP	32/32	32/32	32/32	32/32
CGI	50/50	50/50	50/50	50/50

The second experiment was performed in order to assess the forecasting performances of the proposed approach, namely the capability of predicting the increase or decrease *before* its happening. To this end 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, namely by observing only  $T - k$  signs. It is interesting to notice that the method shows an almost optimal performance in the recognition and forecast of the reliable sequences, and the different risk does not affect the forecast accuracy.

### 3.1 Comparative Analysis

The proposed approach was compared with two other techniques:

- Single Continuous HMM: the main features of the presented approach are the direct modeling of the sign (different than standard modeling of the return) and the employment of two HMMs (different than standard use of one HMM). So a single continuous HMM (Gaussian emission), directly modeling the returns and trained using all sequences, represents the natural competitor of our approach. The recognition / prediction is made by forecasting the next return using the Gaussian of the most probable state given the observed sequence (or forecasting the most probable state in  $k$  time steps).

- Local Polynomial Trend: this method is based on local polynomial trend (LP model – [16]):

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

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 ( $h=2$ ). 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 [8]. 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  $\alpha$  and  $\beta_i$  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 before have been repeated with these models (forecasting with single continuous HMM has been repeated 1000 times in order to ensure statistical significance of the sampling); the results are shown in Table 3 and 4. By comparing the results with Table 2, it could be noted the

**Table 3.** Recognition and forecasting accuracies for different indices (correct/total) when using one single continuous HMM

Index	Recogn. Accuracy	Forecast -1	Forecast -2	Forecast -3
DJ	22.3/26	21.9/26	21.6/26	21.7/26
GE	29.2/30	27.2/30	28.8/30	27.9/30
IBM	34.2/39	26.2/39	33.2/39	29.7/39
BGP	27.3/32	24.6/32	27.1/32	27.1/32
CGI	42.3/50	19.8/50	22.6/50	22.9/50

**Table 4.** Recognition and forecasting accuracies for different indices (correct/total) when using Polynomial Local Trends

Index	Recogn. Accuracy	Forecast -1	Forecast -2	Forecast -3
DJ	25/26	24/26	22/26	20/26
GE	23/30	26/30	23/30	17/30
IBM	36/39	34/39	31/39	24/39
BGP	15/32	21/32	13/32	10/32
CGI	33/50	38/50	29/50	23/50

great increase in accuracy gathered by the HMM modeling strategy proposed in this paper. In particular, for difficult shares (like CGI) the predictions of the other methods are almost completely random.

## 4 Detection of Fluctuant Sequences

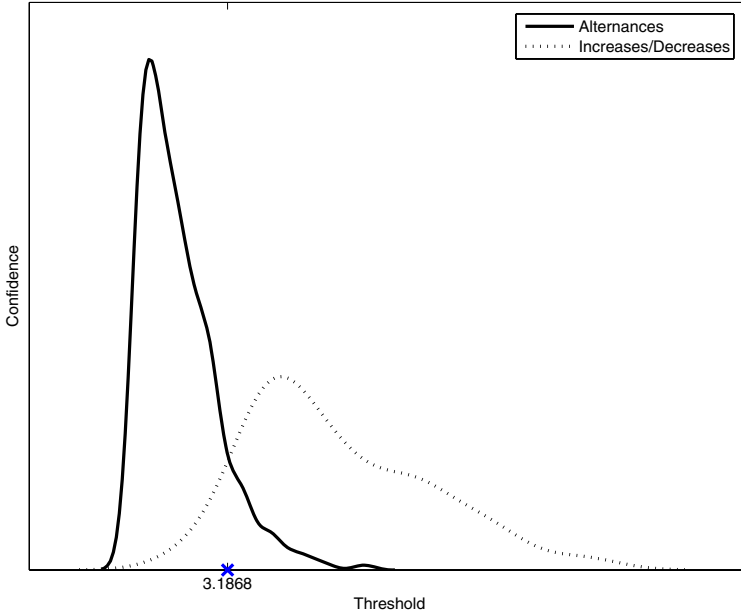
In practical cases, given a generic sequence, the goal of an investor is to get some suggestions for the solution of the following problems: does the sequence present a clear and defined trend? If so, does the sequence represent a reliable increase or a reliable decrease? Further, can we detect such situation in advance? The main purpose of this paper was to answer to the second and third questions. 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 most widely employed rejection rule is the Chow rule [17], 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, which could be easily derived from our asymmetric modeling methodology. In particular the difference between the log likelihoods is considered

$$C(O) = |\log P(O|\lambda_+) - \log P(O|\lambda_-)| \quad (2)$$

If the system does not show a clear evidence in favor of a certain class, i.e.  $C(O) < \theta$  (for a chosen threshold  $\theta$ ), 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  $\theta$  is obviously crucial, and it is linked to the level of confidence desired by the user. In our approach it is directly computed on the basis of the training set: in particular we divide the training set into two subsets, one containing the sequences with a clear trend (i.e. both the sequences of reliable increase and reliable decrease), and one with the others. Then the quantity  $C(O)$  is computed for both the sets and the two empirical pdf's are estimated; the threshold is then computed in a Bayesian way, by considering the point where the two pdfs intersect. We assumed equal priors, even if a different choice based on prior knowledge of the analyzed data set is also possible. The idea is briefly sketched in Fig. 1: the continuous line represents the empirical distribution of the statistic (2) 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  $\theta$  (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  $F_+$  or  $F_-$ .

In Table 5 we show the percentages of the right rejections when trying to recognize all the possible sequences. These accuracies represent the number of



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

**Table 5.** Percentages of correctly rejected sequences

Index	Correct
DJ	93.1%
GE	95.1%
IBM	85.8%
BGP	91.3%
CGI	93.3%

times the system is able to correctly rejecting a sequence belonging to  $F_?$ : we could notice that the percentage is generally high.

## 5 Conclusions

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 exploits the natural asymmetry present in the market by training two separate models, one for the increase situation and one for the decrease. Experiments on different indices show the feasibility of the proposed method.

Future efforts will be devoted to extend the presented model. An immediate attempt will deal with different definitions of local trends (namely what we define as sequences of reliable increase/decrease), for example using some technical



analysis instruments to define the increasing and decreasing periods. Moreover, a more thorough experimental analysis will be made in order to test recognition accuracies on the reliable and unreliable sequences at different rejection levels.

Another interesting research direction is to try to enrich the model by linking 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 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 [18], or modeling 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 [19]), or finally by modeling interactions between couple of processes (as in the Coupled Hidden Markov Models [20]).

## References

1. Campbell, J., Lo, A., MacKinlay, A.: *The Econometrics of Financial Markets*. Princeton University Press, Princeton (1997)
2. IEEE: Special issue on neural networks in financial engineering. *IEEE Transactions on Neural Networks* 12(4) (2001)
3. Cao, L., Tay, F.: Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Trans. on Neural Networks* 14(6), 1506–1518 (2003)
4. Dueker, J.: Markov switching in garch processes and mean-reverting stock market volatility. *Journal of Business and Economic Statistics* 15, 26–34 (1997)
5. Bengio, Y., Lauzon, V.P., Ducharme, R.: Experiments on the application of IOHMMs to model financial returns series. *IEEE Transaction on Neural Networks* 12(1), 113–123 (2001)
6. Liehr, S., Pawelzik, K., Kohlmorgen, J., Lemm, S., Müller, K.R.: Hidden markov mixtures of experts for prediction of non-stationary dynamics. In: *IEEE Proc. of Neural Networks for Signal Processing IX*, pp. 195–204 (1999)
7. Engle, R.: Autoregressive conditional heteroskedasticity with estimates of the variance of u.k. inflation. *Econometrica* 50, 987–1008 (1982)
8. Pring, M.: *Technical Analysis Explained*, 3rd edn. McGraw-Hill, New York (1991)
9. Rabiner, L.: A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proc. of IEEE* 77(2), 257–286 (1989)
10. Christofferson, P., Diebold, F.: Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science* 52(8), 1273–1287 (2006)
11. Hamilton, J.: *Time Series Analysis*. Princeton University Press, Princeton (1994)
12. Akaike, H.: A new look at the statistical model identification. *IEEE Trans. on Automatic Control* AC-19, 716–723 (1974)
13. Frankel, J., Rose, A.: Currency crashes in emerging markets: an empirical treatment. *Journal of International Economics* 41, 351–366 (1996)
14. Holthausen, R., Larcker, D.: The prediction of stock returns using financial statement information. *Journal of Accounting and Economics* 15, 373–412 (1992)
15. Psaradakis, Z.: Spagnolo: On the determination of the number of regimes in Markov-Switching autoregressive models. *Journal of Time Series Analysis* 24, 237–252 (2003)
16. Kendall, G., Stuart, A., Ord, J.: *The advanced Theory of Statistics*. Griffin, London (1983)

17. Chow, C.: An optimum character recognition system using decision functions. IRE Trans. on Electronic Computers 6, 247–254 (1957)
18. Diebold, F., Lee, J., Weinbach, G.: Regime switching with time-varying transition probabilities. In: Hargreaves, P. (ed.) *Nonstationary Time Series Analysis and Cointegration*, pp. 283–302. Oxford University Press, Oxford (1994)
19. Otranto, E.: The multi-chain markov switching model. *Journal of Forecasting* 24(7), 523–537 (2005)
20. Brand, M., Oliver, N., Pentland, A.: Coupled hidden markov models for complex action recognition. In: *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition* (1997)