Prediction of financial time series with Time-Line Hidden Markov Experts and ANNs

GEORGES JABBOUR AND LUCIANO MALDONADO
Escuela de Ingeniería de Sistemas e Instituto de Estadística Aplicada y Computación
Universidad de Los Andes
Núcleo La Hechicera, Edif. B, 2º piso, Ala Sur, Mérida 5101
VENEZUELA
Jabbour@ula.ve, maldonaj@ula.ve http://webdelprofesor.ula.ve/economia/maldonaj

Abstract: - In this article, the use of Time-Line Hidden Markov Experts (THME) in the prediction of financial time series is presented and its efficiency is compared with that obtained using multilayer perceptron neural networks trained with BKP. The THME belongs to a focus known as mixture of experts, whose philosophy consists in decomposing the times series in states. Each expert models a particular state to achieve capturing the time series patterns in a sufficiently precise way, since for every situation in which time series can be found there is one or more experts that have the capacity to generate an adequate prognosis for the given situation. The state transition of each time series is time-variant. Experiments were carried out with 15 series of financial time series in which most of the world's bursatile indexes can be found. The results show that THME models greatly surpass those of Artificial Neural Networks.

Key-Words: - THME, HMM, ANN, financial time series, mixture of experts, fuzzy clustering.

1 Introduction

Economic-financial time series have patterns, which are very difficult to detect. This makes predicting of such time series a very complex task. There are those who maintain that these types of predictions are useless in the short and long term [1]. However, financial analysts and researchers, with innovative points of view, maintain that the possibility of predicting, with a certain level of accuracy, the future behavior of this type of time series using past information exists [2,6]. The point of view presented in this work comes from this same position and is specifically based on the use of THME and ANNs models to predict financial time series.

2 Time-Line Hidden Markov Experts

One of the most recent points of views for complex time series predictions is based on models called, Time-line Hidden Markov Experts (THME), whose philosophy consists in dividing the time series in various states. A state is nothing more than a subset of the series patterns with one of its most important characteristics its homogeneous behaviour, free of chaos and without complex dynamics. Thus, the main idea of the THME models is to train some sub-models in local environments and convert them in experts in their respective environments. They will, later on, be combined to generate a global output that represents the prediction of the THME model.

Under this point of view, for each moment of time, the series is in a particular state exerting influence on its future behavior. Therefore, is necessary to know the way in which the series behaves when it is found in each one of the possible states. This is precisely the objective of each one of the experts. However, this is not enough for the THME model to be able to offer good predictions, given that it is necessary to understand the form in which the referred states evolve in time series. In fact, before predicting the value of the time series, predicting its state is necessary. In this sense, the Hidden Markov Models (HMM) are appropriate for modeling the stochastic process that represents the state of the time series and as such serve to moderate, administrate or control the expert outputs. Nevertheless, a conventional HMM is not capable of describing the transition in each moment of time given that its state transition probability is defined over the complete process. This means, the probabilities of state transition are constant in time, which prevents the HMM from achieving its objective adequately. This will result in inaccurate predictions. In order to resolve such a problem, the matrix of state transitions for the HMM should be variant along the time. This means that instead of using a constant transition matrix A, a transitional matrix A(t) that varies from one moment to another is generated, according to the dynamics of the time series.

With respect to the experts, they can be of any type of connectionist model or regression model. In this work, each expert is a Multilayer Perceptron Neural Network. Moreover, is necessary another connectionist model to predict the probabilities of state transitions. In this case, an RBF network (Neural Network with Radial Base Function) was used. The information generated by this model is required by the HMM for predicting probability that the series is found in each one of the possible states.

To predict the state transitions, the RBF network uses $\Delta X_t = (X_t - X_{t-1})$ as input, where X_t is a vector containing past values of the series. These values are used to predict the value of the serie in time *t*. Therefore, the state transitions are determined by the dynamic situation (or speed) of the series when the prediction is generated.

3 The THME Training

THME training was carried out using the following steps [6]:

3.1 Obtaining the time series states

Time series dynamics for any time t is given by:

$$D_t = \Delta X_t = X_t - X_{t-1} = [(y_{t-1}-y_{t-2}),(y_{t-2}-y_{t-3}),\ldots,(y_{t-L}-y_{t-(L+1)})] \quad (1)$$

M states were built for each time series; applying Fuzzy C-Means Clustering (FCMC) over the patterns of sequential observations according to the equation (1). The result is the matrix $U_{NxC} = \{\mu_{ij}\}$, where μ_{ij} is the degree of membership of the i-th date to the j-th cluster [5,8].

The algorithm obtains the clusters minimizing the following function objective:

$$O = \sum_{t=1}^{T}\sum_{j=1}^{M}\left[\mu_j(t)\right]^2 \left\| D_t - M_j \right\| \quad (2)$$

D is a set of T data. The process ends when function (2) reaches its optimum value. This is, when the matrix U produces the minimum value of this function.

3.2 Experts Training

A multilayered feed-forward neural network carries out the role of expert. Each expert was trained using the Backpropagation algorithm [3,7]. The training patterns were obtained as explained in section 3.1.

3.3 HMM Training

The output of each expert gives the conditional mean of the Gaussian distribution of a HMM state. The HMM was trained with a modified Baum-Welch algorithm based on the EM principle [4,9].

If we assume Gaussian the probability distribution of HMM for the state j, we have that:

$$b_j(y_t) = P(y_t \mid s_t = j, X_t, \lambda') = \frac{1}{\sqrt{2\pi\sigma_j^2}} Exp\left\{-\frac{\left[y_t - \hat{y}_j(X_t)\right]^2}{2\sigma_j^2}\right\} \quad (3)$$

j=1,…,M; t=1,…,T and $s_t \in \zeta$ (ζ is the space of the time series states).

According to this algorithm, the re-estimation of the HMM of each time series was carried out as follows:

$$\tilde{\pi}_i = \frac{P(Y, s_0 = i \mid \lambda')}{P(Y \mid \lambda')}; \text{ i=1,…,M} \quad (4)$$

$$\tilde{a}_{ij}(t) = \frac{P(Y, s_{t-1} = i, s_t = j \mid \lambda')}{P(Y, s_{t-1} = i \mid \lambda')};$$

i=1,…,M; j=1,…,M; t=1,…,T (5)

$$\tilde{\sigma}_i^2 = \frac{\sum_{t=1}^{T}\left[\gamma_i(t)(y_t - \hat{y}(X_t))^2\right]}{\sum_{t=1}^{T}\gamma_i(t)}; \text{ i=1,…,M;} \quad (6)$$

The modified Baum-Welch algorithm guarantees that these re-estimation formulas converge at a local maximum [6].

3.4 State Transition Network

An ANN with radial base function (RBF) was used to predict the probabilities of state transitions, which produces the global output of the model when combined with the outputs of the experts using HMM. Unlike local experts, this network receives ΔX_t instead of X_t.

4 Prediction of Time Series using THME

Given a time serie $Y_{t-1}=(y_1, y_2, ..., y_{t-1})$, the prediction of one step, $\hat{y}_t$, was generated. This was done using the prior probabilities and the posterior probabilities of the states [6]. In the first case (prior prediction), observations of previous moments to t were used. In the second case (posterior prediction), the previous information to t was employed with the ouput from the prior prediction. The prior probability of each state is a coefficient that regulates the output of its expert.

The prediction of the next observation ($\hat{y}_t$) was obtained using the following steps [6]:

(1) Calculation of the prior probabilities of the states:

$$P(s_t = j \mid Y_{t-1}, \lambda') = \sum_{i=1}^{M} a_{ij}(t) P(s_{t-1} = i \mid Y_{t-1}, \lambda') \tag{7}$$

(2) Calculation of the prior prediction:

$$\hat{y}_{pr}(t) = \sum_{i=1}^{M} P(s_t = i \mid Y_{t-1}, \lambda') \hat{y}_i(X_t) \tag{8}$$

(3) Calculation of the posterior probabilities of the states:

$$P(s_t = i \mid Y_t, \lambda') = \frac{P(Y_t \mid s_t = i, \lambda')}{\sum_{j=1}^{M} P(Y_{t-1} \mid s_t = j, \lambda')} = \frac{p(y_t \mid s_t = i, \lambda') P(s_t = i \mid Y_{t-1}, \lambda')}{\sum_{j=1}^{M} p(y_t \mid s_t = i, \lambda') P(s_t = i \mid Y_{t-1}, \lambda')} \tag{9}$$

(4) Calculation of the posterior prediction:

$$\hat{y}_{po}(t) = \sum_{i=1}^{M} P(s_t = i \mid Y_t, \lambda') \hat{y}_i(X_t) \tag{10}$$

Equations (7-10) generate two different predictions, one prior and other posterior. Even though the last one is theoretically more appropriate, there are no guarantees that it will be better. Therefore, upon evaluation of the model, observing both types of predictions to finally select the best one is necessary. Mathematically speaking, this means to equal $\hat{y}_t$ to $\hat{y}_{pr}(t)$ or to $\hat{y}_{po}(t)$.

On the other hand, these expressions offer the mechanism to obtain one step predictions. When multiple step predictions are required, that is, predictions for a time located h units in the future, h predictions of one step should be generated, beginning in the time t and ending in the time t+h–1, so that each obtained prediction is used to generate the prediction of the following time, until reaching the h-th unit of time. This prediction scheme has a more realistic sense than predictions with only one step. For this reason, they were used in this work to carry out the validation of the models.

5 Experiments and Results

The financial time series employed are shown in Table 1.

Name	Origin	No. of Data
IBC	Venezuela	1316
IFC	Venezuela	1315
IIC	Venezuela	1315
BOVESPA	Brazil	1232
CAC40	France	1264
DAX	Germany	1265
DJ	USA	1256
FTSE100	England	1260
NASDAQ	USA	1256
NIKKEI225	Japan	1229
NYSE	USA	1256
SEOUL	Korea	1222
SHANGHAI	Hong Kong	1200
SP500	USA	1256
SM	Switzerland	1195

Table 1. Time Series under Study

The data in these series corresponds to the daily closing value of each financial index. The study period is different for each series. In addition, the series was analyzed in two different ways. One based on the daily frequency data and the other based on their weekly averages. This was carried out with the objective of reducing the duration of training, given the quantity of time series and given the number of models to evaluate for each series. Moreover, it was a way of eliminating part of the noise present in each serie.

Only the IBC, IFC and IIC series were analyzed daily while the complete series were used on a weekly frequency. On the other hand, 85% of each series was used for training (from left to right) and 15% for validation.

In Tables 2 and 3, the best THME models obtained are shown. The parameters used for each model and for each series were selected in a pilot experiment: (a) Markov Hidden Model: 0.01 as a training error and 1000 maximum cycles. (b) State transition network: a training error of 0.01. (c) ANN Experts: 10^{-5} as training error and 2000 maximum cycle.

Time Series	Number of Experts	Nodes per Expert	MSE (A Priori)	MSE (A Posteriori)
IBC	10	5	89980	***75080***
IFC	5	20	***193870***	196520
IIC	12	20	***221830***	224340
BOVESPA	10	20	***598470***	642320
CAC40	7	10	222230	***217350***
DAX	7	10	***181060***	185750
DJ	5	8	19494	***18729***
FTSE100	7	10	120780	***119170***
NASDAQ	10	20	36003	***34052***
NIKKEI225	6	8	***71380***	74900
NYSE	10	10	65000	***63001***
SEOUL	10	10	***17962***	20870
SHANGHAI	5	5	***2344***	3112
SP500	10	8	1124	***1049***
SM	7	10	189270	***174600***

Table 2. THME predictions (weekly frequency)

Time Series	Number of Experts	Nodes per Expert	MSE (A Priori)	MSE (A Posteriori)
IBC	10	5	127100	***125950***
IFC	10	4	***130450***	133370
IIC	10	5	***205950***	209740

Table 3. THME predictions (daily frequency)

The last two columns of Tables 2 and 3 show the mean squared error (MSE) of the prior and the posterior prediction, respectively. During the validation process, the predictions with the lowest error was chosen. On the other hand, it is important to highlight that the THME dimensions, in terms of the number of parameters, is still similar when some series have more data than others. To reach the best models, an average of 11,72 models per serie were evaluated.

However, in order to evaluate the quality of the predictions generated by the THME models, predictions for the same time series were obtained using pure multilayered perceptron feed-forward ANNs trained with BKP. The stop criteria for training each ANN was the same, a training error equal to 10^{-5} and a maximum number of cycles equal to 10000. It depended on which occurred first, without provoking overtraining. The best results obtained are shown in Tables 4 and 5. Reaching these models was done by averagely evaluating 10,67 ANNs for each series.

Time Series	Number of Inputs	Nodes Layer Hidden 1	Nodes Layer Hidden 2	MSE
IBC	1	5	5	***126670***
IFC	1	5	0	***203190***
IIC	1	7	0	***446640***
BOVESPA	2	10	10	***2158400***
CAC40	1	10	0	***1228200***
DAX	4	1	0	***189940***
DJ	2	10	0	***19598***
FTSE100	2	10	0	***120800***
NASDAQ	3	5	0	***21848***
NIKKEI225	3	3	0	***76350***
NYSE	1	20	0	***70310***
SEOUL	2	10	0	***29217***
SHANGHAI	3	10	0	***12666***
SP500	1	15	0	***1278***
SM	1	10	0	***463090***

Table 4. ANN predictions (weekly frequency)

Time Series	Number of Inputs	Nodes Layer Hidden 1	Nodes Layer Hidden 2	MSE
IBC	3	8	4	***124920***
IFC	3	20	0	***178330***
IIC	1	10	5	***421130***

Table 5. ANN predictions (daily frequency)

Tables 6 and 7 show the MSE of the best ANN and the best THME for each one of the series of weekly and daily frequencies. In these tables, the values of column A come from (MSE_ANN/MSE_THME) *100%. These values show the percentage relationship between ANNs and the THMEs, which is necessary to point out, due to the fact that MSE changes substantially from one time series to another (while the greater the average in the time series, the greater the MSE).

Therefore, a universal error measurement is needed for all the series. This will show how greater or how lesser (in percentage) is the error generated by an ANN compared to the one generated by a THME model.

Time Series	MSE ANN	MSE THME	A
IBC	126670	75080	168,713372%
IFC	203190	193870	104,807345%
IIC	446640	153960	290,101325%
BOVESPA	2158400	598470	360,652998%
CAC40	1228200	217350	565,079365%
DAX	189940	181060	104,904452%
DJ	19598	18729	104,639863%
FTSE100	120800	119170	101,367794%
NASDAQ	21848	34052	64,1606954%
NIKKEI225	76350	71380	106,962735%
NYSE	70310	63001	111,601403%
SEOUL	29217	17962	162,66006%
SHANGHAI	12666	2344	540,358362%
SP500	1278	1049	121,830315%
SM	463090	174600	265,229095%

Table 6. Comparison between THME and ANN (weekly frequency)

Time Series	MSE ANN	MSE THME	A
IBC	124920	125950	99,1822%
IFC	178330	130450	136,7037%
IIC	421130	205950	204,4817%

Table 7. Comparison between THME and ANN (daily frequency)

For the weekly frequency data, we can observe, in 14 of the 15 time series, that the THME models had a better performance. In 6 of those cases the difference oscillated between 1.36% and 11.50%, and in two more the MSE of the ANNs surpassed that of the THME models in ranges between 20% and 68%. In the rest of the series, the ANN error surpassed the THME error by at least 68% and the difference reached 465% in some cases. This indicates that the THME models generate predictions with considerably greater accuracy than the ANNs.

For data of daily frequency, for the IBC time serie, the result was that the MSE of the THME model was greater than the MSE obtained with the ANN, but with a difference lesser than 1%. This means, for the IBC time serie, they had a very similar accuracy. However, for the IFC and IIC time series the MSE of the ANNs was greater than those of the THME in 136,7% and 204,5%, respectively. This ratifies the superiority of the THME models in the prediction of financial time series.

To conclude, the Figures 1-4 show as an example, the predictions of the DAX and CAC40 series, where the behavior of the THME and the ANNs can be seen.

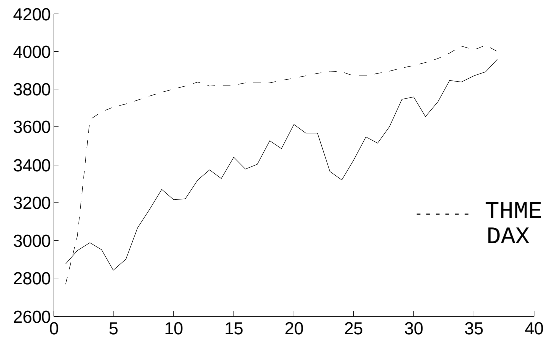

Fig 1. THME Prediction of the DAX serie

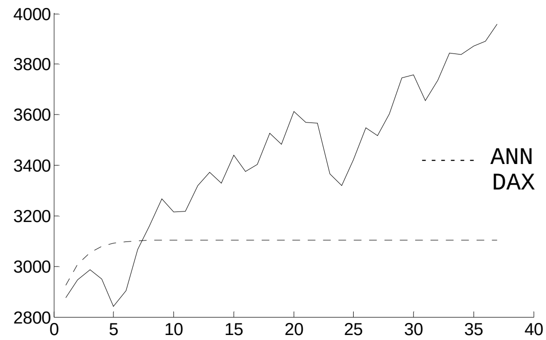

Fig 2. ANN Prediction of DAX serie

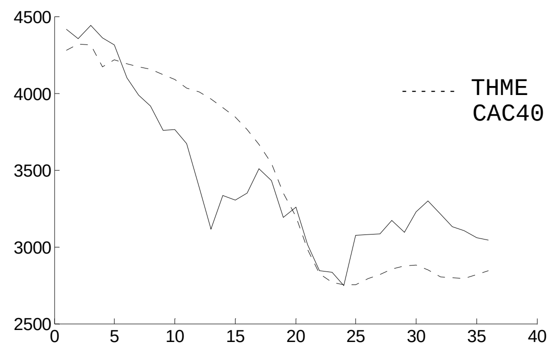

Fig 3. THME Prediction of the CAC40 serie

Fig 4. ANN Prediction of the CAC40 serie

These graphs show that the predictions carried out by the THME models not only surpass those of the ANN in accuracy, but its tendencies are similar to the original time series in a greater measure.

It is worth mentioning that such a similarity is independent from the accuracy of the method. For example, for the DAX serie, the accuracy of both methods was similar (MSE_THME = 189940; MSE_ANN = 181060). However, the prediction of the tendencies is significantly better in the case of THME. On the other hand, in the case of the CAC40, the prediction capacity of the THME model was much greater. This shows that the THME models posses an outstanding capacity to predict financial time series not only for accuracy but also because they were able to capture the patterns in an effective manner.

6 Conclusions

Regarding the accuracy of predictions, the conclusion is that the THME models produce better results than the ANNs. We can also add the fact that THME models have a greater capacity to capture the patterns of time series. It would be interesting to use another regression model instead of ANNs, in order to improve the results obtained in this research. For example, Support Vector Machines could be used as experts [10].

The THME models have the particularity that its topological structure does not necessarily have to be modified when passing from a time series to another. This means, two time series of different sizes can be modelled with two THMEs that have a similar number of parameters, such as the case of the daily and weekly time series where even though the daily frequency series had approximately four times more data than the weekly frequency series, the structure of the THMEs were similar.

The THME models have the disadvantage of requiring a lot of experimentation and computing time to achieve the best model. On the other hand, a greater expertise is required for a THME model to be developed. In this sense, future works could be oriented to face this problem, using an optimization method like Genetic Algorithms, where the chromosomes would represent the model structure, with the MSE of the prediction as the evaluation function.

The use of ANNs in the prediction of financial time series cannot be discarded due to the fact they require less computing costs than THMEs to obtain better models, indeed, for ANNs, there are a lot of procedures for model structure optimization [11].

7 Acknowledgment

The authors wish to thank the referees for their comments and suggestions. This work was supported by the CDCHT-Universidad de Los Andes, under Grant E-252-06-02-B and I-1047-07-02-C.

References:

[1] Burton, M. *A Random Walk Down Wall Street.* Norton,1996.

[2] Chorafas, D. *Chaos Theory in the Financial Markets.* Irwin, 1994.

[3] Gupta, M.; Jin, L. and Homma, N. *Static And Dynamic Neural Networks.* John Wiley And Sons, 2003.

[4] Rabiner, L. *A Tutorial on Hidden Markov Models*. IEEE, Vol. 77, No.2. 1989.

[5] Jang, J.; Sun, C. y Mizutani, E. *Neuro-Fuzzy and Soft Computing.* Prentice Hall, 1997.

[6] Wang, X.; Whirgham, P. and Deng, D. Time-Line Hidden Markov Experts and its Application in Time Series Prediction. *The information science, discussion paper series.* No. 2003/03 june 2003. ISSN 1172-6026, Otago University, 2003.

[7] Witten, I.; Frank E. *Data Mining Practical Machine Learning Tools and Techniques*. Elsevier, 2005.

[8] The MathWorks. *Fuzzy Logic Toolbox User's Guide*. The MathWorks, Inc., 2006.

[9] Bulla, J.; Bulla , I. Stylized facts of financial time series and hidden semi-Markov models. *Computational Statistics & Data Analysis*. 51 (2006) 2192 – 2209. Elsevier, 2006.

[10] Papadimitriou, S.; Terzidis, K. Prediction and Dynamical Reconstruction of non-stationary data with Delay-Coordinates Embedding and Support Vector Machine Regression. *4th WSEAS Int. Conf. on Non-linear Analysis, Non-linear Systems and Chaos*. Sofia, Bulgaria. 60-67, October 2005.

[11] Sureerattanan, S.; Phien, H.; Sureerattanan, N.; Mastorakis, N. The Optimal Multi-layer Structure of Backpropagation Networks. *Proceedings of the 7th WSEAS International Conference on Neural Networks*. Cavtat, Croatia. 108-113, June 2006.