Predicting stock markets in boundary conditions with local models

Gianluca Bontempi, Edy Bertolissi, Mauro Birattari
Iridia - Université Libre de Bruxelles

Av. F. Roosevelt 50, CP 194/6
1050 Bruxelles - Belgium
http://iridia.ulb.ac.be
e-mail: {gbonte,eberto,mbiro}@ulb.ac.be

1 Introduction

Financial time series present many distinctive features. An important one is the fact that large positive and large negative observations (outliers) occur more frequently that in other economic time series and tend to appear in clusters. This fact is well-known in financial literature that devoted a wide amount of work to this issue. In particular, two are the main research directions which addressed this problem in the past: the modeling of the volatility and the theory of extreme values.

The first approach, *modeling of the volatility*, exploits the fact that outliers emerge in clusters in order to construct a model for the outliers themselves. However, since in empirical finance it is assumed that the evolution of the asset prices is better described by random walks, the aim of this approach is to model the variance rather than the level of the variable to predicted. It is also said that these models are used to forecast volatility, that is the condition of the market when the level changes rapidly. An example of these models are the so-called Autoregressive Conditional Heteroskedasticity (ARCH), initially proposed in Engle [8]. ARCH model had a large number of applications in forecasting problems concerning exchange rates, stock markets, interest rates, aggregate and individual indices, and so on.

The second approach, *extreme value theory* [7], focuses on the prediction of models for extreme values of variables. The idea is that traditional parametric statistical and econometric models are ill-suited to the treatment of extreme events. Parametric models aim to return a good fit in regions where most of the data fall, at the expense of good fit in the tails where, by definition, few observations fall. The idea of extreme value theory is to fit models using only the extreme event data rather than all the data, thereby fitting the tail distributions, and only the tail distributions.

The two approaches share the common idea that a certain amount of regularity exists in the boundary regions and that this regularity can be exploited in order to build accurate models. However, while the first approach deals mainly with a prediction of the variance, the second addresses the problem of density estimation.

This paper adopts the idea of regularity in the boundaries of financial time series in order to fit forecasting models which are able to outperform random walk predictions. In particular we propose the adoption of a local learning technique, called Lazy Learning, in order to perform model estimation and prediction in extreme conditions.

The Lazy Learning technique has been tested on a series of difficult modeling and control problem, including the participation to the International Competition on Time Series in Leuven [11], where the Lazy technique ranked second and the participation to the Third International Erudit competition[1] where the Lazy technique was awarded as a runner-up.

In this paper the Lazy Learning method is proposed to return predictions in extreme conditions of the trend of the Italian stock market index. The experiments show that in boundary conditions the technique is able to outperform a random predictor and to return a significant rate of accuracy.

2 Prediction in boundary conditions

Given the time series s^t, consider the series r^t, obtained by taking the difference between the logarithms of adjacent values of s^t:

$$r^t = \log s^t - \log s^{t-1} = \log \frac{s^t}{s^{t-1}} \qquad (1)$$

[1] http://www.erudit.de/erudit/activities/ic-99/index.htm

Most approaches merely try to predict the next value of the time series r^t. Such point prediction is appropriate only if the signal-noise ratio is favorable. Unfortunately, this is not the case for financial time series where the noise is often larger than the signal itself.

An alternative could be the prediction of the trend $y_h^t = \text{sign}(s^{t+h} - s^t)$, of the time series s^t over the next h steps. Given the monotonicity of the logarithm, y_h^t can be expressed in terms of the series r^t:

$$\begin{aligned} y_h^t &= \text{sign}\big(\log s^{t+h} - \log s^t\big) \\ &= \text{sign}\Big(\sum_{i=1}^{h} r(t+i)\Big) \end{aligned} \tag{2}$$

This task however remains a formidable prediction problem and there is little evidence in literature that a predictor of the quantity 2 could improve the performance of the simplest random predictor.

This paper addresses the problem of predicting the quantity 2 in boundary conditions. The idea originates from that area of probability and statistics known as *extreme value theory* [7] and consists in fitting a model to observed data focusing only on the extreme event data rather than on all data.

We start to describe our method by briefly introducing some notation. Let express the trend as

$$y_h^t = \text{sign}(\Delta_h^t) \tag{3}$$

where $\Delta_h^t = \log s^{t+h} - \log s^t$.

Suppose that the trend can be described by a probabilistic model

$$P(y_h^t | x^t) \tag{4}$$

where $x^t = \{r^t, r^{t-2}, \ldots, r^{t-n-1}\}$ is given by the previous n values of the series r^t.

The prediction of the trend y_h^t on the basis of the past n values of r^t, can be obtained from an estimation of the quantity

$$E[\Delta_h^t | x^t] = F(x^t) \tag{5}$$

where E denotes the expectation operator and F is a generic nonlinear function of the regressor x^t.

An extreme value version of the prediction problem 5 consists in performing the prediction only at those values of t for which a given boundary condition is met. This condition is expressed in the form $v(t) > Th$, where $v(t)$ is the normalized rate

$$v(t) = \frac{r(t)}{\sum_{i=1}^{t} r(i)^2} \tag{6}$$

and Th represents a threshold value.

For the values of t for which the boundary condition is satisfied the problem 5 can be expressed as the estimation of the quantity

$$E[\Delta_h^t | x^t, v(t) > Th] \tag{7}$$

This paper is concerned with the estimation of the quantity 7 on the basis of historical observations. Based on the extreme value theory, we propose an approach to the problem of forecasting the trend of a financial time series on the basis of historical data. So far, we have discussed only the definition of the prediction task, without referring to any particular technique for the approximation of the function F of equation 5. The next section will introduce our local learning technique for prediction in boundary conditions.

3 Local modeling approximator

Lazy learning [1] is a local learning technique which postpones all the computation until an explicit request for a prediction is received. The request is fulfilled interpolating locally the examples considered relevant according to a distance measure. Each prediction requires therefore a local modeling procedure that can be seen as composed of a *structural* and of a *parametric* identification. The parametric identification consists in the optimization of the parameters of the local approximator. The structural identification involves, among other things, the selection of a family of local approximators, the selection of a metric to evaluate which examples are more relevant, and the selection of the *bandwidth* which indicates the size of the region in which the data are correctly modeled by members of the chosen family of approximators. For a comprehensive tutorial on local learning and for further references see Atkeson *et al.* [2].

In previous published works [5, 6, 4], the authors proposed a method in which the family of local approximators and the bandwidth are selected locally and tailored for each query point by means of a local leave-one-out cross-validation [10]. The problem of bandwidth selection is then reduced to the selection of the number k of neighboring examples which are given a non-zero weight in the local modeling procedure. The *Lazy Learning Toolbox for use with Matlab*© [3] implements a Lazy Learning technique, which considers polynomials of different degrees and allows a local model selection, as well as a local combination of approximators of different degrees. In particular, each time a prediction is required for a specific query point, a set of local models is identified, each with a different polynomial degree and each including a different

number of neighbors. The generalization ability of each model is then assessed through a local leave-one-out. Finally, a prediction is obtained either combining or selecting the different models on the basis of some statistic of their cross-validation errors.

The major feature of this toolbox consists in the adoption of the *recursive least squares* algorithm for the identification of the local models [4]. This is an appealing and efficient solution to the intrinsically incremental problem of identifying and validating a sequence of local linear models centered in the query point, each including a growing number of neighbors. It is worth noticing here that a leave-one-out cross-validation of each model considered does not involve any significant computational overload, since it is obtained though the PRESS statistic [9] which simply uses partial results returned by the recursive least squares algorithm. This technique is now implemented in the Lazy Learning Toolbox for use with Matlab© publicly available on the Web[2].

4 Experimental results

To assess the reliability of the predictive model proposed in the previous section, we conducted a series of experiments on real financial series. Results are presented for the Italian stock market index (MIBTEL). We used the data from 10 Jan 1994 to 14 Mar 1997 as the training set and the data from 15 Mar 1997 to 3 Dec 1999 as the *out-of-sample* test set.

This section reports as series of experiments where the Lazy Learning technique was adopted to predict the trend $y_h^t = \text{sign}(\Delta_h^t)$ for different regressor orders n, different horizons h, and different thresholds Th. The figures illustrate the percentage of correct predictions as a function of the different parameters. Note that a value greater then 50% indicates that our prediction method outperforms a generic random predictor. While the predictive accuracy of the model with low threshold does not improve the simplest random model, the prediction in boundary conditions (thus with sufficiently large thresholds) shows a clear improvement. Figure 1 depicts the accuracy as a function of the order n of the regressors for a fixed horizon $h = 5$ and a fixed threshold $Th = 1.3$. It is possible to see that the accuracy of the prediction changes with the order. In particular the best results, which are close to a 70% accuracy, are obtained when the order of the regressor is chosen equal to 13. Regressor orders equal to 9 and 17 produce as well good results. Once the order has been identified, the influence of the value of the threshold Th on the quality of the predictions

[2]http://iridia.ulb.ac.be/~lazy/

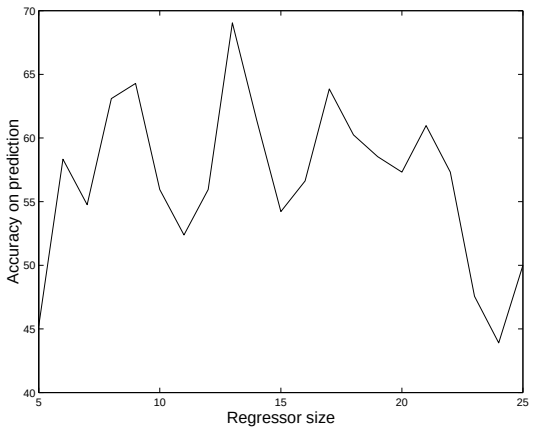

Figure 1: Accuracy on prediction as a function of the order of the regressor using $Th = 1.3$.

is analyzed in figure 2. The top three graphs show the accuracy on prediction as a function of Th, while the bottom graph indicates the number of predictions performed, i.e. the number of times the boundary condition is met. For $Th = 0$, the prediction is always performed. The number of the prediction performed decreases as the threshold increases. It is possible to see in all the graphs that the prediction accuracy is larger for values of the threshold between 1 and 2. This shows that the accuracy of the prediction is increased if the prediction is performed only when the boundary condition is met. In particular, for $n = 13$ it is always possible to achieve a prediction accuracy higher the 67% for Th between 1.2 and 1.9. Figure 3 shows the accuracy on prediction as a function of the horizon, for $n = 13$ and $Th = 1.5$. In this case it is possible to see that the best accuracy on prediction is achieved for an horizon equal to 5 days, i.e. for one trading week. Figure 4 shows a three dimensional plot of the accuracy on prediction as a function of the horizon h and the threshold Th using an order of the regressors $n = 13$. This plot shows a ridge in correspondence of $h = 5$, which highlights that the one trading week horizon yields the best accuracy for the values of Th inside the range of interest. Moreover, figure 4 allows a comparison between the unconditional prediction and the approach we propose in this paper for extreme value prediction. The accuracy of the former is represented by the values assumed on the line $Th = 0$, while a higher accuracy can be obtained by the extreme value prediction method for values of the threshold in the range $1 < Th < 2$.

The approach proposed in the paper attains an accuracy larger than 65% on the prediction of the trend and therefore can be considered as a valid alternative to random walk prediction.

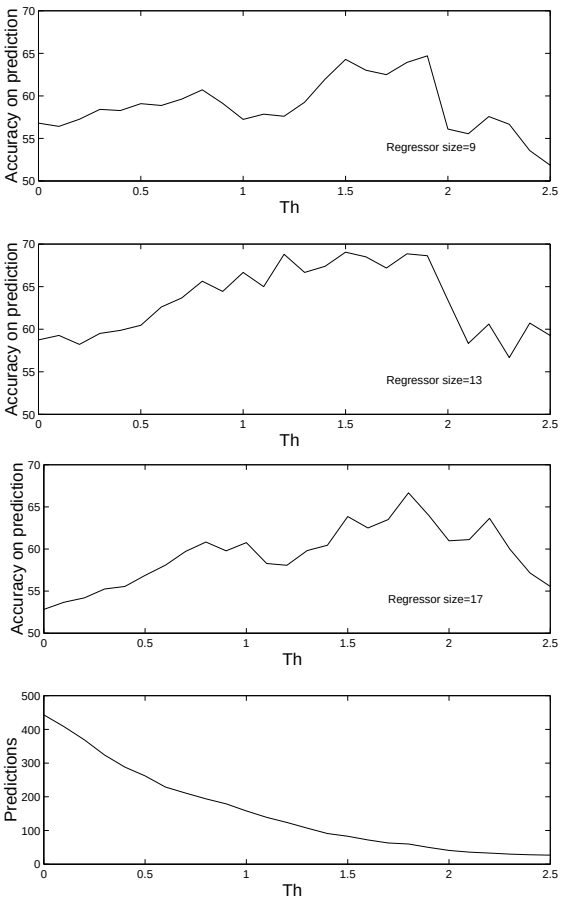

Figure 2: Accuracy on prediction as a function of the threshold, for regressor order of 9, 13 and 17. The bottom graph indicated the number of predictions performed.

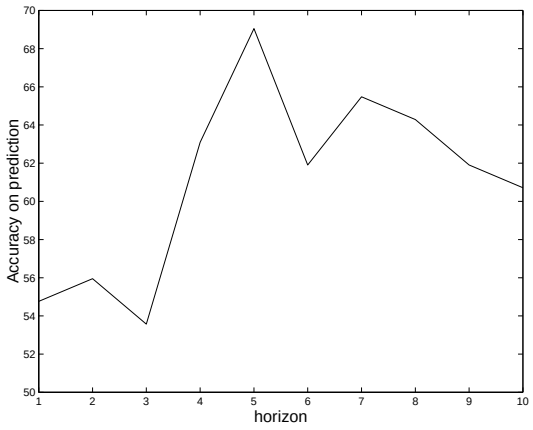

Figure 3: Accuracy on prediction as a function of the horizon. For $Th = 1.5$, and $n = 13$.

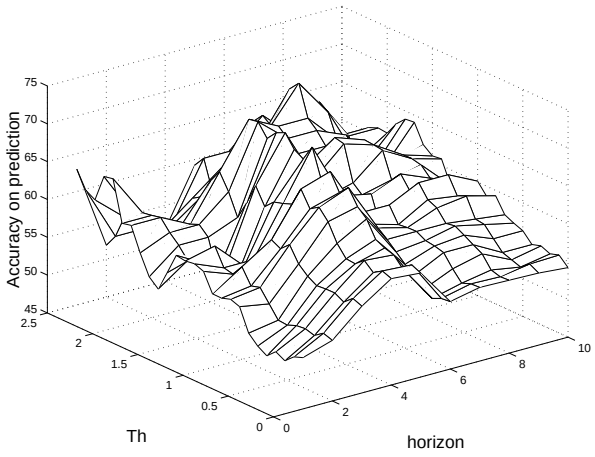

Figure 4: Accuracy on prediction as a function of the horizon and the threshold. Regressor order $n = 13$.

References

[1] D. W. Aha. Editorial. *Artificial Intelligence Review*, 11(1–5):1–6, 1997.

[2] C. G. Atkeson, A. W. Moore, and S. Schaal. Locally weighted learning. *Artificial Intelligence Review*, 11(1–5):11–73, 1997.

[3] M. Birattari and G. Bontempi. The lazy learning toolbox, for use with matlab. Technical Report TR/IRIDIA/99-7, IRIDIA-ULB, Brussels, Belgium, 1999.

[4] M. Birattari, G. Bontempi, and H. Bersini. Lazy learning meets the recursive least-squares algorithm. In M. S. Kearns, S. A. Solla, and D. A. Cohn, editors, *Advances in Neural Information Processing Systems 11*, pages 375–381, Cambridge, 1999. MIT Press.

[5] G. Bontempi, M. Birattari, and H. Bersini. Recursive lazy learning for modeling and control. In *Machine Learning: ECML-98 (10th European Conference on Machine Learning)*, pages 292–303, 1998.

[6] G. Bontempi, M. Birattari, and H. Bersini. Lazy learning for modeling and control design. *International Journal of Control*, 72(7/8):643–658, 1999.

[7] P. Embrechts, C. Klupperberg, and T. Mikosch. *Modelling Extreme Events*. Springer, 1997.

[8] R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of the U.K. inflation. *Econometrica*, 50:987–1008, 1982.

[9] R. H. Myers. *Classical and Modern Regression with Applications*. PWS-KENT Publishing Company, Boston, MA, second edition edition, 1994.

[10] M. Stone. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society B*, 36(1):111–147, 1974.

[11] J. A. K. Suykens and J. Vandewalle, editors. *Nonlinear Modeling: Advanced Black-Box Techniques*, chapter The K. U. Leuven Time Series Prediction Competition, pages 241–251. Kluwer Academic Publishers, 1998.