Error-Based Learning Processes on Financial Markets

DIMA BOGDAN[a], PIRTEA MARILEN GABRIEL[a], CRISTEA ŞTEFANA[b]
a. Finance Department, Faculty of Economics and Business Administration
a., b. West University of Timişoara
J. H. Pestalozzi nr. 16, 300115, Timişoara
ROMANIA
bogdandima2001@gmail.com http://www.feaa.uvt.ro

Abstract: - A critical issue in financial markets' research is the debate between *the academic orthodox approach* of the *Efficient Markets' Hypothesis* and the critics rising from the *behavioral finance* paradigm and practice. Several alternative explanations have been proposed in order to provide a more realistic description of the financial markets' inner mechanisms. This debate's importance consists in the implications of the adopted point of view on the assessment of the financial markets' predictability degree. [4], [5] proposed a unified approach labeled as Adaptive *Markets Hypothesis*. Thus, from a practical point of view, the main issue consists in providing a pertinent answer to the next question: Is an active portfolio management able to provide better results that a passive "follow the market on long-run" strategy? If in *adaptive models* the markets are considered to display, at least in a certain sense, a degree of predictability, then, one of the major difficulties in supplying empirical evidences is the requirement of knowing ex ante the "exact" forecast model used by the economic subjects. The aim of this study is providing a solution to this problem inspired by the *transduction*'s (supervised learning) algorithms. Our main output consists in the thesis that the forecasting errors matter for price formation in financial markets. So, despite the fact that nor the theoretical foundations nor the empirical evidences are conclusive, we argue that the nature of the "exact" learning mechanisms can be seen as one of the key variables in investors' decisions and markets evolution. There is a significant positive payoff of a more detailed study of such mechanisms inside an extended framework of financial markets as complex systems.

Key-Words: - Financial markets, *FTSE 100*, *Adaptive Markets Hypothesis*, forecasting algorithms, forecasting errors, adaptive mechanisms

1 Introduction

Nowadays, one of the most important issues in financial markets' analysis is the divergence between two major conceptual frameworks: the *Efficient Markets Hypothesis* - assuming that markets fully, accurately, and instantaneously incorporate all available information into market prices and the *behavioral finance* which tries to account for/ considers the behavioral idiosyncrasies of the markets participants. However, Lo [2004, 2005] proposes a new conceptual framework which reconciles these theories and implies that the markets are efficient with behavioral alternatives, by applying the principles of evolution - competition, adaptation, and natural selection - to financial interactions. This framework was labeled as *Adaptive Markets Hypothesis*.

As Lo [2005,1] states "Based on evolutionary principles, the Adaptive Markets Hypothesis implies that the degree of market efficiency is related to environmental factors characterizing market ecology such as the number of competitors in the market, the magnitude of profit opportunities available, and the adaptability of the market participants".

If such a framework is viable, that it implies the postulate of the existence of several adapting learning mechanisms on financial markets. Or, in other words, it presumes that the economic subjects are able to learn from their forecasting mistakes according to the accuracy, relevance, volume and structure of the market information. However, there are important difficulties in order to empirically test this hypothesis, since such a test requires knowing ex ante the "exact" forecasting model used by the market participants. The purpose of our study is to advance a possible solution to this problem by employing a two-stage approach: (1) a competitive framework of identifying some plausible forecasting algorithms and (2) an ex post test of their forecasting error relevance in the formation of observed level of prices.

2 The analytical framework

The canonical model of the *Efficient Markets Hypothesis* underlines the assumption that market agents systematically behave in a "rational" fashion by taking optimal decisions based on a benefits-costs trade-off in a "correct" market framework. This conventional wisdom was recently challenged by a set of empirical evidences which records the existence of different types of specific behavioral biases in the market agents' decisional processes that leads to "sub-optimal" outputs. Several alternative explanations have been proposed in order to provide a more realistic description of the financial markets' inner mechanisms. Among them, itcan be noticed the *Adaptive Markets Hypothesis* or the *Swing Market Hypothesis* (Pan [2003]). Such explanations shared some common views:

- There are different relevant time frames, informational asymmetries and operational capacities for the markets operators (see for related issues and empirical evidences Cheong et al [2006]);
- There are different institutional structures on markets (different investors "species");
- The markets display areas of informational efficiency alternating with inefficiency zones and there are some switching mechanisms between these two stages;
- The markets movements are driven by different types of forces and the global status is the output of their combination;
- There are several psychological determinants of investors' decisions such as greed, fear, over-confidence and overreaction, loss aversion, herding, mis-calibration of probabilities, regrets, asymmetric attitude towards risk etc. As a consequence, the investment decisional processes are not purely "rational" but rather "rational and emotional" ones depending on various factors able to influence the investors' attitude towards the decisional variables ("Emotions are the basis for a reward-and-punishment system that facilitates the selection of advantageous behavior"- Lo [2005, 27]).

One of the most important consequences of such views consists in the thesis that financial markets are, at least for some periods and in certain conditions, predictable. In fact, from a practical point of view, the main issue consists in providing a pertinent answer to the next question: Is an active portfolio management able to provide better results that a passive "follow the market on long-run" strategy? If more information could lead to better results, than there is an incentive for investors to engage themselves in the search and use of a supplementary amount of information in order to obtain an "informational rent". As is argued by Black [1986] such rent is provided by "noise traders"- investors who are trading based on spurious information which is merely noise. Of course, a greater amount of information does not necessarily imply a higher capacity to provide more accurate forecasts. To illustrate this idea, let's consider the evolution of the predictive power of a simple moving average rule according to the length of the rolling window used in its construction. More exactly, for the close values of the *FTSE 100* index between April 1984 and April 2009, such a rule could be defined as:

$$\begin{aligned} &if\ Close_{t-1} > MA_{t-1}(w)\ then\ Close_t < Close_{t-1} \\ &and \\ &if\ Close_{t-1} < MA_{t-1}(w)\ then\ Close_t > Close_{t-1} \end{aligned} \quad (1)$$

The appliance of the rule for values of w will generate the results reported in Table 1. Table 1: The predictive accuracy of the *MA* rule (number of correct identified cases*100/ number of total cases)

w	**Up**	**Down**
5	47.48	52.06
25	42.47	57.85
50	38.96	61.39
75	36.22	64.18
100	34.91	65.17
125	35.52	64.39
150	34.72	64.20
175	33.51	64.96
200	32.56	66.30
225	31.77	66.15
250	31.20	66.60
275	30.18	67.63
300	29.66	67.87
325	29.56	67.91
350	29.13	68.29

It can be noticed that an increase in the window length generates an asymmetric impact on the rule predictive accuracy: it decrease the number of cases correctly identified for the upward cases and it increases the number of cases founded in downward trends. Even such a simple example could show that the critical variable is not only the volume of information but rather its accuracy, relevance, completeness and usefulness. In addition, it could provide intriguing evidence in advancing some hypothesis about the asymmetric use of information in upward / downward market movements. Finally,

it suggests that the issue of information structure per se should be complemented with a fair model of market operators' mechanisms of acquiring, collecting, using and distributing this information. A feasible approach to such model should *inter alia* consider the structure of the market. More exactly, it can be observed that the modern capital markets are dominated by institutional investors not by individual ones. The investments funds (with a whole range of structures - from mutual, closed-end, unit investment or exchange traded to real estate or sovereign investment) and banks, investors' collective organisms and semi-public entities are the major market-makers both at national and international level. All this structures are using an important amount of financial and time resources to collect and to interpret an enormous volume of information by engaging professional R&D and information processing components. As a consequence, what matters is not only the individual investor's psychology but also the organizational culture and decisional infrastructure of the institutional entities. Even more, for that type of investors the markets are creating special incentive mechanisms to stimulate the distribution of information. As Donaldson [2003] has argued: "Under the current system, distributions of market data revenues to self-regulatory organizations are based primarily on each self-regulatory organization's reported trade volume. This compensation scheme has created a financial incentive for self-regulatory organizations to report as many trades as possible". If this is the case, then, the position of the *behavioral finance* should be extended in order to account for the effects of specific organizational cultures and institutional network interactions.

Our starting point is based on the next hypothesis:

- Like the *Adaptive Markets Hypothesis* we emphasize the fact that the capital markets are heterogeneous with different types of investors behavior and only partially informational efficient;
- For individual and institutional investors as well the decisions are partially adopted in a rational manner and partially under the impact of emotional and organizational factors;
- There are incentives for investors to acquire a higher volume of relevant information, although they are not all the time capable to accurately and fully process the entire set of gathered data;
- Still, the investors are able to learn from their mistakes and to correct their forecasts;
- The adjustment speed of wrong predictions is not identical for each investor, so that, at a observable market scale, there will not be a predominance of "one period" adjustment processes.
- As a consequence, the evolution of financial assets' prices is partially predictable by using the current and historical information.

Formally, in this framework the price forecast y^*_t could be described as:

$$y^*_t = F_t + \varepsilon_t \qquad (2)$$

Here F_t is the "rational" part of the forecasting mechanisms which incorporate the objective information and ε_t is the subjective component which encapsulate both psychological and organizational factors as well as the effects of informational asymmetries (the subjective "noise").

Since the relevant information cover both historical and current periods and the noise is not necessarily just a simple stochastic variable, then:

$$y^*_t = \omega_{i=t-k\ldots t-1}(F_i) + I_t + \\ + \theta_{i=t-k\ldots t-1}(\varepsilon_i) + E_t \qquad (3)$$

ω, θ denote the time-depending forecasting algorithms while *I, E* are the current sets of objective and subjective forecasting elements.

A critical aspect of such approach concerns the fact that both individuals and organizations are able to learn from their experiences so their expectations are time-depending variables. Of course, in order to address this issue, it is necessary to identify the learning mechanisms and their effects on anticipation formation process. Although the classical views deal with abstraction, data compression, simplification and summarization of operational elements of learning, the modern psychology advances different models of this process: *rule-based* theories, *simple* and *multiple prototype* theory, the so-called *exemplar* theories (such as ALCOVE), *explanation-based* theories or *Bayesian* approaches (see for instance Ratcliff [2006], Feldman [2000, 2003], Rendell [1986], Brown [1994]). Despite the significant differences between them, all these approaches share a common wisdom:

- In the learning process, there is a *core* structure (a rule, a "central tendency", a specific instance or a "prototype") which serves as main input;
- In achieving a certain output of learning, the subjects are "clustering" the input data by trying

to create several operational decomposition mental structures;
- The learning mechanisms involve a set of probabilities of observing the given data, on the assumption that it was generated by a certain abstract structure;
- There are mutual interdependencies between the level of performance expected from the learner and the types of contents of the material to be learned;
- All the learning practices consist in series of discrete mental operations connected to the rules, examples, recall, practice, prerequisites, objectives, helps, mnemonics, and feedback;
- Data that display consistent patterns may be compressed without losing any important information (of course, this last statement could be also seen as a pious desire that the *Ockham's razor* principle applies, *entia non sunt multiplicanda praeter necessitatem* and the learner can obtain a "simple enough" output model to be manageable; otherwise, the sceptical point of view that "how to deal with complexity is still a complex problem" or more simple "was the good Franciscan friar really use a razor?" still makes a point).

If such wisdom is considered, then the prices' forecast formal model could be rewritten as:

$$y^*_t = \left[\omega^{core}{}_{i=t-k\ldots t-1}(F_i) + \theta^{core}{}_{i=t-k\ldots t-1}(\varepsilon_i)\right] + \\ + (I_t + E_t) + \xi_t \qquad (4)$$

Here ω^{core}, θ^{core} denote the "persistent" data patterns and subjective probabilities attributed to the fitting concepts while ζ labels various types of "shocks".

In the meantime, there could be done a cut-off between the individual and institutional investors' learning mechanisms. More exactly, it could be considered that the institutional investors have the possibility to involve technical and fundamental analysis methods in their market analysis more often that the individual ones. Even more, they can employ *supervised learning* procedures (such as Neural Network-Multi-layer Perceptron, Support Vector Machines, k-Nearest Neighbours, Gaussian Mixture Model, Gaussian, Naive Bayes, Decision Tree or RBF classifiers) in order to identify key patterns in data movements. These procedures assume that the data set is an independent and identically-distributed random variables' sample drawn according to an unknown probability distribution. If the *iid hypothesis* does not hold, than these procedures' output usually fails inside *50-50 accuracy* result. However, if such a cut-off is taken into account, then the general prices forecasting model becomes:

$$y^*_t = \left[\begin{matrix}\omega^{machine\_core}{}_{i=t-k\ldots t-1}(F_i) \otimes \omega^{human\_core}{}_{i=t-k\ldots t-1}(F_i) + \\ + \theta^{core}{}_{i=t-k\ldots t-1}(\varepsilon_i)\end{matrix}\right] + \\ + (I_t + E_t) + \xi_t \qquad (5)$$

In this general framework, in order to identify a feasible solution for the error - based learning process a simply competitive approach could be involved. Such an approach could imply the next steps:

1. The *ex ante* selection of a forecasting algorithms' set $\{F_i\}$;
2. The estimation of the "optimal-fitting" parameters on a learning window with length l of observed data so the estimation errors for the current level of the forecasted variable y_t, ε_t, are minimal regardless their sign:

$$\varepsilon_t = |y_t - F_t| \rightarrow Min \qquad (6)$$

3. The computation of each individual algorithm' forecasting performance with these parameters for q steps ahead according with a pre-defined performance criterion;
4. The competitive elimination of a certain z percent of the algorithms for which the relative forecasting performance falls bellow the established trash in order to
5. Develop the "second generation" of algorithms as linear or non-linear combinations of the "first generation' surviving ones.

Of course, these steps should be repeated by developing and eliminating new generations of algorithms until a certain generation k is selected according to a predefined "saturation" criterion.

The purpose is to declare a winner – the algorithm with the best forecasting capacity, to estimate the forecasting errors inside a "learning area" according to this algorithm and to test the existence of certain adaptive mechanisms in the prices formation in respect to these errors.

Suppose that two algorithms are considered:

$$F1: y^*_t = \sum_{i=t-u}^{t} \frac{j}{\sum_{j=1}^{u} j} y_i + \sum_{s=t-m}^{t} w_s \sigma^2{}_s$$

$$with \sum_{s=t-m}^{t} w_s = 1 \qquad (7)$$

$$F2: y^*_t = \delta_t e^{\delta_t} \sum_{s=t-n}^{t} \mu_s y_s \qquad (8)$$

The first algorithm is a linear combination between a weighted moving average of observed predicted variable's values and its *historical volatility* defined as a convex combination of the standard deviation's lagged values.
The second one is the linear product between a time-varying parameter, its exponential value and an average of the predicted value.
The first algorithm is designated to smoothly adjust the central tendency of the forecasted variable by its short / long run observed volatility, while the second one captures the extreme values of the observational data.
For the first algorithm, it could be noticed that, if a small number of w_s parameters are considered, then a "short time" form of volatility correction is involved. Conversely, if a large number of such parameters are considered then the "long-run" data volatility corrects the central tendency.
For the second algorithm, a particular case of interest appears when $n = 0$, $w_1 = -1$, $\delta_1 = \delta_2 = \ldots = \delta_t = -h$ and the algorithm collapses to:

$$F2: y^*_t = he^{-h} y_t \qquad (9)$$

This version has a shape for certain values of the h parameter close to a *Poisson distribution* and, like this, can be applied to data with a large number of possible events, each of which is rare.

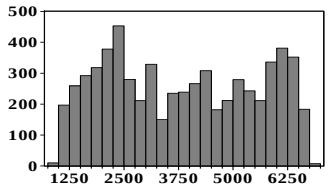
3 The empirical evidences

An important aspect in these algorithms' implementation concerns the data frequency. More exactly the number of turning points tends to increase with the shift from low to high frequency data, so the behavior of the algorithms' parameters must be sensitive to the chosen data frequency.
Thus, one of the critical issues in the analysis of movements in financial assets' prices is generated by the fact that if prices' changes are independent, there should not be any noticeable streaks in the data. Or, in fact, the empirical evidence shows that such increasing and decreasing streaks are highly frequent - in a manner that is improbable under the classic Gaussian model. An alternative approach was proposed by Mandelbrot, Fisher and Calvet [1997] with the so-called Multifractal Model of Asset Returns (MMAR) and largely discussed and developed in the literature (Mandelbrot and Hudson [2004]; Eisler and Kertész [2004]; Lux [2003]; Lux and Kaizoji [2004]). The meta-assumption of this approach consists in the thesis that the dynamics of prices' evolution reflects a fractal property with the same characteristics as the initial data series. This property stays intact when shifting from low to high "resolution" (on the time-scale). Inserting a random component to this property guarantees a more accurate description of the prices' behavior. Of course, this implies that the economic subjects are neutral in respect to the informational leverage time scale (i.e. they are acting in the same way no matter what is the frequency of the new information relevant to their decisions). If this assumption does not hold, then the time-invariance property of data is questionable and these data' characteristics will significantly vary over different time frequencies.
Consider, for instance, the evolution of the *FTSE 100* index between April 1984 and April 2009. *FTSE 100* is a share index of the 100 most highly capitalized UK companies listed on the London Stock Exchange. FTSE 100 companies represent about 81% of the market capitalization of the whole London Stock Exchange. Even though, the *FTSE All-Share Index* is more comprehensive, the *FTSE 100* is by far the most widely used UK stock market indicator. For daily and, respectively, monthly data the main statistic characteristics of the index have evolved as in Fig.1 (source of data: http://finance.yahoo.com/).
Apparently, the data are displaying long right "fat tails" effects and a distribution that is *flat* (platykurtic) relative to the normal one for both time frequencies. However, the Jaque-Bera test reveals significant differences between the two time scales. Jarque-Bera is a test statistic for testing whether the series is normally distributed.
The statistic test measures the difference of the skewness and kurtosis of the series with those from the normal distribution. The statistic is computed as:

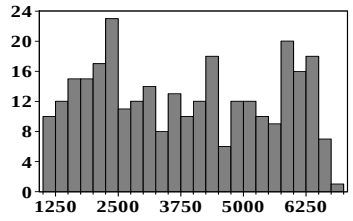
$$Jarque-Bera = \frac{N}{6}\left(S^2 + \frac{(k-3)^2}{4} \right) \qquad (10)$$

where S is the skewness distributional parameter and K is the kurtosis one.

Fig.1 The distribution characteristics of the *FTSE 100* index



Daily data	
Observations	6312
Mean	3837.175
Median	3776.300
Maximum	6930.200
Minimum	986.9000
Standard Deviation	1693.115
Skewness	0.075238
Kurtosis	1.677684
Jarque-Bera	465.8158
Probability	0.000000



Monthly data	
Observations	301
Mean	3838.146
Median	3759.300
Maximum	6930.200
Minimum	1009.400
Standard Deviation	1694.976
Skewness	0.074146
Kurtosis	1.682993
Jarque-Bera	22.02943
Probability	0.000016

Under the null hypothesis of a normal distribution, the Jarque-Bera statistics is distributed as χ^2 with 2 degrees of freedom. The reported *Probability* is the probability that a Jarque-Bera statistics exceeds (in absolute value) the observed value under the null hypothesis - a small probability value leads to the rejection of the null hypothesis of a normal distribution. In this example, there are important differences between the test values for the two considered time scales as well as between their associated probabilities. Of course, this could not be seen as a proof per se, but rather as a caveat about the limits of the time-invariance hypothesis. A direct consequence of this consists in the fact that the predictive capacity of an individual forecasting algorithm could vary across different time scales.
Furthermore, we consider the daily frequency of the *FTSE 100* data. The empirical distribution tests reject the hypothesis of a normal distribution for this data set (Table 2).

Table 2: Empirical distribution tests for *FTSE 100* daily data

Hypothesis: Normal
Included observations: 6323

Method	Value	Adj. Value	Probability
Lilliefors (D)	0.097342	NA	0.0000
Cramer-von Mises (W2)	16.79878	16.80011	0.0000
Watson (U2)	16.73237	16.73369	0.0000
Anderson-Darling (A2)	113.4043	113.4177	0.0000

Method: Maximum Likelihood – degree of freedom corrected (Exact Solution)

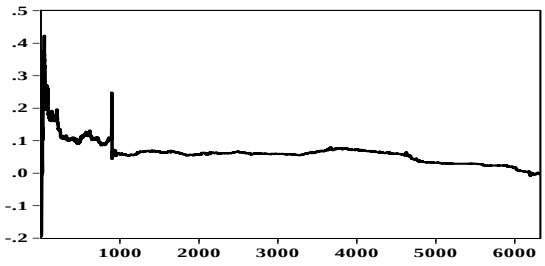
Parameter	Value	Std. Error	z-Statistic	Prob.
MU	3832.425	21.32196	179.7408	0.0000
SIGMA	1695.464	15.07809	112.4455	0.0000

Table 3: The random walk test for *FTSE 100* daily data

Method: Maximum likelihood (Marquardt)

	Final State	Root MSE	z-Statistic	Prob.
ε_t	3983.700	48.68967	81.81818	0.0000

Fig.2 The *AR(1)* parameter of the *FTSE 100*



The fitting of the parameters will be done at each observation in order to minimize the current fitting error.
The pre-defined and the averages values of the estimated parameters are reported in Table 4. The *F1* algorithm was considered a combination between a weighted moving average of the close values of the *FTSE 100* and a 5 days historical volatility, while *F2* was estimated as a product between a simple moving average of the close levels and a time varying factor. The common idea is to capture the "short-run" trend adjustments in index levels.

Table 4: The pre-defined and estimated (average) values of parameters

F1:	
u	5
m	5
w_1^{est}	0.29
w_2^{est}	0.15
w_3^{est}	0.15
w_4^{est}	0.16
w_5^{est}	0.25
F2:	
n	5
$\mu_1=\mu_2=\mu_3=\mu_4=\mu_5$	0.20
δ^{est}	0.57

The involved estimation procedure is based on the so-call *Generalized Reduced Gradient* method (GRG). This method implies corresponding adjustments of the variables, so that the active constraints continue to be satisfied as the procedure moves from one data to another. The pre-determined and the estimated values of the parameters are used

to derive the one step-ahead forecast and to estimate the corresponding errors. The main characteristics of these errors look like in Table 5.

Table 5: The general properties of the one step-ahead forecast errors

Included observations: 6312							
	Mean	Median	Max	Min.	Quant. *	Skew.	Kurt.
-30, -20)	-22.35	-22.35	-22.35	-22.35	-22.35	NA	NA
[-20, -10)	-12.63	-11.67	-10.15	-17.20	-11.67	-0.99	2.45
[-10, 0)	-1.32	-0.97	0.00	-9.80	-0.97	-2.32	11.12
[0, 10)	0.66	0.50	5.59	0.00	0.50	2.24	11.09
All	*-0.73*	*-0.51*	*5.59*	*-22.35*	*-0.51*	*-2.28*	*19.17*

Included observations: 6312							
	Mean	Median	Max	Min.	Quant.	Skew.	Kurt.
[-30, -20)	-11.08	-11.08	-11.08	-11.08	-11.08	NA	NA
[-20, -10)	-6.49	-6.54	-5.07	-8.08	-6.54	-0.07	1.51
[-10, 0)	-0.83	-0.64	0.00	-4.87	-0.64	-1.90	8.03
[0, 10)	0.84	0.65	4.99	0.00	0.65	1.87	7.85
All	*6.29*	*5.78*	*9.68*	*5.05*	*5.78*	*1.20*	*3.53*

Autocorrelation	**Partial Correlation**		**AC**	**PAC**	**Q-Stat**	**Prob**
|***** |	|***** |	1.00	0.62	0.62	2441.50	0.00
|*** |	| |	2.00	0.36	-0.04	3274.00	0.00
|** |	| |	3.00	0.20	-0.01	3533.80	0.00
|* |	|* |	4.00	0.17	0.10	3722.50	0.00
|* |	| |	5.00	0.13	-0.02	3823.10	0.00

Autocorrelation	**Partial Correlation**		**AC**	**PAC**	**Q-Stat**	**Prob**
| |	| |	1.00	-0.03	-0.03	5.12	0.02
*| |	*| |	2.00	-0.07	-0.07	33.41	0.00
*| |	*| |	3.00	-0.17	-0.18	215.31	0.00
| |	*| |	4.00	-0.06	-0.08	236.29	0.00
**| |	**| |	5.00	-0.21	-0.25	509.99	0.00

The errors display the characteristics of a non-normal distribution with some long left fat tail effects in the case of first algorithm's errors and long right fat tail effects for the second algorithm's errors and "flat" (*platykurtic*) peakedness relative to the normal one for both sets of errors. The same output results for the specific empirical distribution tests for both series of errors. It is important to note that the distributional parameters are not conserved over the errors data subsets. Also, the tabulation suggests that there are some variations over time in both sets of errors with some important extreme values. The correlogram of both errors' sets suggests that up to lag 1, there can be some significant autocorrelations on their level. Even more, the *BDS Independence portmanteau test* (with bootstraps) indicates that the data are not adequately described by the *independent and identical distributed hypothesis* (Table 6). Such results raise the question of the possibility to manipulate the errors data set inside a pattern recognition algorithm and could impose the appeal of local adaptive parameters for *F1* and *F2*. There is nothing surprising in that the distribution of errors change over time from "trend" to "oscillation" market areas.

Table 6: The *BDS Independence portmanteau test* for errors (included observations: 6312)

Dimension	BDS Statistic	Std. Error	z-Statistic	Normal Prob.	Bootstrap Prob.
2	0.062576	0.001111	56.31722	0.0000	0.0000
3	0.101557	0.001761	57.65688	0.0000	0.0000
4	0.123920	0.002092	59.22621	0.0000	0.0000
5	0.134830	0.002175	61.97692	0.0000	0.0000
6	0.137773	0.002093	65.82855	0.0000	0.0000
Raw epsilon	1.826818				
Pairs within epsilon	27993384	V-Statistic	0.702621		
Triples within epsilon	1.35E+11	V-Statistic	0.537812		

Dimension	BDS Statistic	Std. Error	z-Statistic	Normal Prob.	Bootstrap Prob.
2	0.020871	0.001079	19.34983	0.0000	0.0000
3	0.042703	0.001714	24.90853	0.0000	0.0000
4	0.059936	0.002042	29.35476	0.0000	0.0000
5	0.070782	0.002128	33.25633	0.0000	0.0000
6	0.076679	0.002053	37.35253	0.0000	0.0000
Raw epsilon	1.554715				
Pairs within epsilon	28076082	V-Statistic	0.704697		
Triples within epsilon	1.36E+11	V-Statistic	0.539443		

Supplementary, it could be noticed that the *Quandt-Andrews Breakpoint Test* of errors' levels (Table 7) rejects the null hypothesis of no structural changes having a maximum perturbation attained in the first half of March 2002.

Table 7: Quandt-Andrews unknown breakpoint test

Quandt-Andrews unknown breakpoint test
Null Hypothesis: No breakpoints within trimmed data
Equation Sample: 2 6312
Test Sample: 948 5364
Number of breaks compared: 4417

Statistic	**Value**	**Prob.**
Maximum LR F-statistic (Obs. 4616)	8.52	0.05
Exp LR F-statistic	1.85	0.06

Ave LR F-statistic	3.14	0.04

Number of breaks compared: 4417		
Statistic	**Value**	**Prob.**
Maximum LR F-statistic (Obs. 4619)	35.71	0.00
Exp LR F-statistic	11.44	0.00
Ave LR F-statistic	9.08	0.00

Note: Probabilities calculated using Hansen's (1997) method

In the same time, the evidences reject the hypothesis that the errors could be modelled as random walk (with drift) processes (Table 5). This hypothesis is tested like:

$$error_t = \mu + error_{t-1} + \varepsilon_t \qquad (11)$$

where μ is the drift parameter and $\varepsilon_t \sim N(0,\sigma^2)$ is a "white noise". This result is connected to the existence of some high-order autocorrelations in the errors.

Table 8: The random walk tests for errors

	Final State	**Root MSE**	**z-Statistic**	**Prob.**
ε	*0.00*	*1.33*	*0.00*	*1.00*
Log likelihood	-10760.58	**Akaike info criterion**		3.41
Parameters	2.00	**Schwarz criterion**		3.41
Diffuse priors	0.00	**Hannan-Quinn criterion**		3.41

	Final State	**Root MSE**	**z-Statistic**	**Prob.**
ε	*0.00*	*1.77*	*0.00*	*1.00*
Log likelihood	-12553.41	**Akaike info criterion**		3.98
Parameters	2.00	**Schwarz criterion**		3.98
Diffuse priors	0.00	**Hannan-Quinn criterion**		3.98

Summing up, the errors appear to be non-normallly distributed, with a distributional structure changing over time and not fairly described by random walk processes. This raises the question of the competitive elimination mechanism. At this point, it can be argued that, since their characteristics are not uniform, a single elimination criterion is not enough to accurately discriminate between them. Thus, a list of such criteria should be involved. Let us consider, for instance, such a set of criteria resumed in Table 9.

Table 9: The elimination criteria

	Average / Standard Deviation less than 1	Distribution close to the normal one (Jarque-Bera statistic less than 10)	Independent and identically distributed	No AR(k)	Fairly described by a random walk process
F1:	Pass	Fail	Fail	Fail	Fail
F2:	Pass	Fail	Fail	Fail	Fail

It can be observed that both algorithms pass only the first criteria of a low average and volatility, but fail to fulfil the others. In such a case, supplementary criterion must be considered in order to select the surviving algorithm. For instance, if the issue of stationarity is taken into account, then the *Kwiatkowski-Phillips-Schmidt-Shin* reveals some differences between error series:

Table 10: *Kwiatkowski-Phillips-Schmidt-Shin* stationarity test

			LM-Stat.
F1:			**0.178892**
F2:			**0.082782**
Asymptotic critical values*:		1% level	0.216000
		5% level	0.146000
		10% level	0.119000

However, for a confidence level of 1%, the test rejects the null hypothesis of second algorithm's forecasting errors' stationarity, but confirms this hypothesis for the first one. Thus, this supplementary criterion could be used to discriminate between them. Furthermore, if the first algorithm is selected, then in order to test its capacity to generate forecasting errors - that are incorporated in market learning mechanisms - a binary variable could be constructed and a *logit regression* model could be estimate. The binary variable takes the value of "1" if the absolute level of errors is less that a certain threshold (in this example 3%) and "0" otherwise. In order to distinguish between the influence of the level and of the errors' sign, such a model could incorporate as separate variables the positive and negative errors generated by *F1* (for an example on the relevance of such a classical model in classifying input vectors in economic models, see for instance Sarlija et al [2006,156]). In such model, the distribution for the errors looks like:

$$\Pr\left(y_i = 1 \middle| \mathbf{x}_{1,2}\right) = 1 - \left(\frac{e^{-x'_{1,2}\beta}}{\left(1+e^{-x'_{1,2}\beta}\right)}\right) = \frac{e^{x'_{1,2}\beta}}{\left(1+e^{x'_{1,2}\beta}\right)} \qquad (12)$$

Table 11: The binary regression statistics

Method: ML - Binary Logit (Quadratic hill climbing)
Included observations: 6311
Covariance matrix computed using second derivatives

	Coefficient	Std. Error	z-Statistic	Prob.
$Errors^+$	0.720912	0.061889	11.64845	0.0000
$Errors^-$	0.116860	0.016589	7.044472	0.0000
Mean dependent variable	0.596894		S.D. dependent variable	0.490561

S.E. of regression	0.490970	Akaike info criterion	1.352536
Sum squared residuals	1520.794	Schwarz criterion	1.354675
Log likelihood	-4265.927	Hannan-Quinn criterion	1.353277
Avg. log likelihood	-0.675951		
Obs with Dep=0	2544	Total observations	6311
Obs with Dep=1	3767		

Goodness-of-Fit Evaluation for Binary Specification
Andrews and Hosmer-Lemeshow Tests
Grouping based upon predicted risk (randomize ties)

	Quantile of Risk			Dep=0		Dep=1	T.	H-L
	Low	High	A.	Expect	A.	Expect	Obs	Value
1	0.0684	0.4303	443	384.577	188	246.423	631	22.7265
2	0.4303	0.4532	374	351.553	257	279.447	631	3.23628
3	0.4532	0.4668	289	340.514	342	290.486	631	16.9288
4	0.4668	0.4767	249	333.279	382	297.721	631	45.1703
5	0.4767	0.4851	237	327.534	394	303.466	631	52.0339
6	0.4851	0.4928	215	322.422	416	308.578	631	73.1859
7	0.4928	0.5037	192	317.576	439	313.424	631	99.9683
8	0.5038	0.5570	168	296.879	463	334.121	631	105.660
9	0.5572	0.6330	177	257.381	454	373.619	631	42.3967
0	0.6331	0.9826	200	180.266	432	451.734	632	3.02249
		T.	***254 4***	***3111.98***	***3767***	***3199.02***	***6311***	***464.329***

H-L Statistic	464.3291	**Prob. Chi-Sq(8)**	0.0000
Andrews Statistic	486.5683	**Prob. Chi-Sq(10)**	0.0000

Note: Total= T; Actual = A

Based on the results of this model, it could be drawn the so-called *Expectation-Prediction (Classification) Table* by involving the one lag forecasting errors. This displays the correct and incorrect classification based on a user specified prediction rule and on expected value calculations. Each observation is classified as having a predicted probability that lies above or below a cutoff p= 0.5 value (Table 12).
The table and associated statistics depict the classification results based upon the expected value calculations. “Correct” classifications are obtained when the predicted probability is less than or equal to the cutoff and the observed y=0, or when the predicted probability is greater than the cutoff and the observed y=1. In our results, 51.28 percentage points of the Dependent = 0 observations and 52.02 percentage points of the Dependent = 1 observations are correctly classified by the estimated model.

Table 12: The *Expectation-Prediction Table for F1 algorithm*

	Estimated Equation			***Constant Probability***		
	Dep=0	**Dep=1**	**Total**	**Dep=0**	**Dep=1**	**Total**
E (of Dep = 0)	1304.57	1807.41	3111.98	1025.50	1518.50	2544.00
E (of Dep = 1)	1239.43	1959.59	3199.02	1518.50	2248.50	3767.00
Total	2544.00	3767.00	6311.00	2544.00	3767.00	6311.00
Correct	1304.57	1959.59	3264.17	1025.50	2248.50	3274.00
% Correct	***51.28***	***52.02***	51.72	40.31	59.69	51.88
% Incorrect	48.72	47.98	48.28	59.69	40.31	48.12
Total Gain*	10.97	-7.67	***-0.16***			
Percent Gain**	18.38	-19.03	-0.32			

**Change in "% Correct" from default (constant probability) specification; **Percent of incorrect (default) prediction corrected by equation*

The fraction of y=1 observations that are correctly predicted is termed the *sensitivity*, while the fraction of y=0 observations that are correctly predicted is known as *specificity*. Overall, our model correctly predicts 51.72% of the observations.
The gain in the number of correct predictions obtained in moving from the right of Table 12 to the left of the table provides a measure of the predictive ability of your model. The gain measures are reported in both, absolute percentage increases (**Total Gain**) and as a percentage of the incorrect classifications in the constant probability model (**Percent Gain**). The estimated model improves on the Dependent = 0 predictions by 10.97 percentage points, but does more poorly on the Dependent = 1 predictions (-7.67 percentage points). Overall, the estimated equation is -0.16 percentage points - worst at predicting responses than the constant probability model.
Finally, one crucial aspect of this methodology concerns the robustness of the involved algorithms. Suppose for instance that the *n* parameter of *F2* is modified from “5” to “24”. Now, the errors generated by this algorithm still do not pass the criteria of distribution close to the normal one, independent and identical distributed, no auto-regressive behavior up to a certain lag *k* and random walk evolution; but they become stationary (Table 13).
In such case, the stationarity could not longer be used in order to discriminate among the algorithms and no surviving one could be chosen. A solution to this kind of situation consists in involving into the learning mechanism a combination of the forecasting errors generated by the two algorithms, so that the weights of such combination are minimizing the absolute level of the resulted error.

Table 13: *Kwiatkowski-Phillips-Schmidt-Shin* stationarity test for modified *F2* parameter errors

			LM-Stat.
F2:			**13.50253**
Asymptotic critical values*:		1% level	0.216000
		5% level	0.146000
		10% level	0.119000

For such a combination, the errors continue to fail the initial set of criteria, although they are stationary. The *logit regression* of empirical parameters looks for these errors as follows:

Table 14: The binary regression statistics for the combined errors

Method: ML - Binary Logit (Quadratic hill climbing)
Included observations: 6296
Covariance matrix computed using second derivatives

	Coefficient	Std. Error	z-Statistic	Prob.
$Errors^+$	9.378997	0.174980	53.60036	0.0000
$Errors^-$	-4.226752	0.055023	-76.81771	0.0000
Mean dependent variable	0.985070	S.D. dependent variable		0.121283
S.E. of regression	0.318843	Akaike info criterion		0.756978
Sum squared residuals	639.8551	Schwarz criterion		0.759122
Log likelihood	-2380.967	Hannan-Quinn criterion		0.757721
Avg. log likelihood	-0.378171			
Obs with Dep=0	94	Total observations		6296
Obs with Dep=1	6202			

It could be observed that the number of binary variables labelled as "0" (cases in which the error is greater than 3%) had substantially decreased. Such evolution explains both the relevant improvement in the prediction of cases for which Dependent = 1 at 77.98 percentage points - the overall correct prediction of 77.08 percentage points- as well as that this new model improves on the Dependent = 0 predictions by 16.01 percentage points, but loses as contributory explanation on Dependent = 1 predictions with 20.52 percentage points comparative to the constant probability specification (Table 15). Meantime, this model could be used to estimate the marginal effect of explanatory variables on the conditional probability.

Table 15: The *Expectation-Prediction Table for combined errors*

	Estimated Equation			***Constant Probability***		
	Dep=0	**Dep=1**	**Total**	**Dep=0**	**Dep=1**	**Total**
E (of Dep=0)	16.45	1365.39	1381.84	1.40	92.60	94.00
E (of Dep=1)	77.55	4836.61	4914.16	92.60	6109.40	6202.00
Total	94.00	6202.00	6296.00	94.00	6202.00	6296.00
Correct	16.45	4836.61	4853.06	1.40	6109.40	6110.81
% Correct	***17.50***	***77.98***	***77.08***	1.49	98.51	97.06
% Incorrect	82.50	22.02	22.92	98.51	1.49	2.94
Total Gain*	16.01	-20.52	-19.98			
Percent Gain**	16.25	-1374.56	***-679.15***			

Change in "% Correct" from default (constant probability) specification;Percent of incorrect (default) prediction corrected by equation*

This marginal effect is given by:

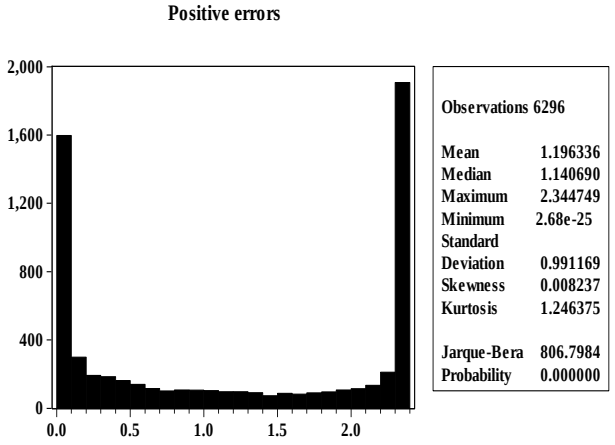
$$\frac{\partial E\left(y_i \mid x_i, \beta\right)}{\partial x_{ij}} = f\left(-x_i^{'}\beta\right)\beta_j \qquad (13)$$

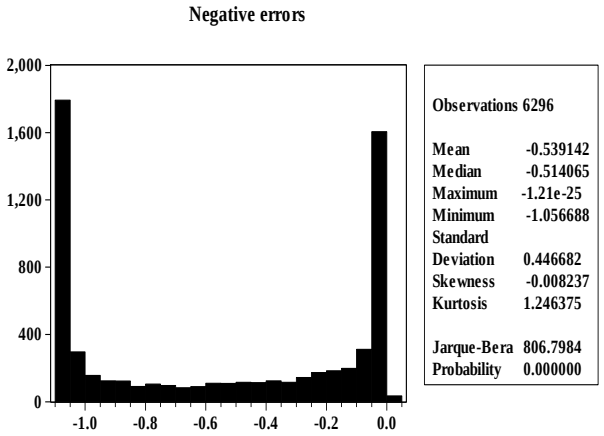
Here $f\left(x\right)=\frac{dF\left(x\right)}{dx}$ is the density function corresponding to F. It should be noted that β_j is weighted by a factor f that depends on the values of all of the regressors in x. The direction of the effect of a change in x_j depends only on the sign of the β_j coefficient. More exactly, positive values of this coefficient imply that increasing x_j will increase the probability of the response; negative imply the opposite. Further, an alternative interpretation of the coefficients results from noticing that coefficients' ratios provide a measure of the relative changes in the probabilities in the sense that:

$$\frac{\beta_j}{\beta_k} = \frac{\dfrac{\partial E\left(y_i \mid x_i, \beta\right)}{\partial x_{ij}}}{\dfrac{\partial E\left(y_i \mid x_i, \beta\right)}{\partial x_{ik}}} \qquad (14)$$

The main statistic characteristics of the positive and negative errors' marginal contributions are reported in Figure 3. It results that these contributions have a non-normal distribution and a relative important deviation comparing to their mean level.

Fig.3 The distribution characteristics of the errors' marginal contributions





Such an example illustrates the importance of a preamble robustness check of the algorithms involved in the learning mechanism. Of course, this stylised example is in practice complicated by the fact that the *robustness areas* of different algorithms are interlinked and the optimization procedures should consider *all* the involved parameters. If such a fitting is not possible, then the derivation of a feasible combination between the surviving algorithms could be not possible.

4 Conclusion

This study investigates the reaction of the financial markets to the forecasting errors as a part of a learning process in the conditions of informational non-efficiency. The main idea is that if the markets are displaying different types of informational imperfections, the economic subjects are forced to learn from their forecasting errors. The meta-hypothesis beyond such idea is that financial markets can be best described as *complex adaptive systems*. If this is the case, then it could be admitted that these markets display areas of local informational efficiency in combination with non-efficient functioning of intrinsic mechanisms moments with some switching between these two phases. In such framework, we conclude that institutionally heterogeneous, bounded rational and based on imperfect information's decisions of market operators could be partially predicted. The main issue in empirically proving this consists in the difficulty to identify the "correct" forecasting model. Our solution is to competitively eliminating a set of predefined forecasting algorithms, to combine the surviving ones and to test *ex post* the relevance of the previous forecasting errors. The results are mixed ones, suggesting that there is a certain relevance of a volatility trend correction algorithm but this is far from being the "exact" one. Supplementary, it seems that a weighted combination of these two algorithms provides better results. However, the viability of such a conclusion is affected by the limits of the study.

Among them we shall refer to:

[1] the limited number of considered forecasting algorithms;

[2] the absence of a more detailed argumentation of the selection criteria;

[3] the single market character of the study. For instance, there are solid evidences that the major European stock markets are cointegrated and each of them is acting as short time leading indicators for the others - Filis and Costas [2006];

[4] the limited exploration of the *adaptive market hypothesis*' consequences etc.

Overall, it appears that even such a limited study reveals some crucial aspects of describing a formal learning mechanism. Among them:

- The importance of a initial robustness check for the involved algorithms;
- The sensitivity of the output to the selection criteria;
- The non-trivial identification of the forecasting errors' main characteristics;
- The local stable nature of the forecasting algorithms' parameters;
- The importance of the number of steps considered in the competitive elimination of algorithms.

Still, even if these aspects are considered, it remains as a major issue the *ad hoc* nature of the forecasting algorithms - more exactly, the distinction between the individual and institutional investors. Thus, the question is: How well do the mentioned algorithms - and more -specific, each algorithm involved - fit the way in which those *species* of investors are formulating *de facto* their forecasts? In order to

address such issues there are two solutions: 1) to verify *ex post* the descriptive accuracy of algorithms as it is done in this study and, respectively, 2) to examine *ex ante* the fundamentals of different types of investors' approaches. And it is clear that there are no prior reasons to discriminate between the results of these two ways to deal with the artificial nature of the considered forecasting methodologies.

Nevertheless, beyond all these caveats, the main output consists in the thesis that the forecasting errors are relevant for financial markets' evolution. More generally, despite the fact that nor the theoretical foundations nor the empirical evidences are conclusive, we argue that the nature of the "exact" learning mechanisms could be seen as one of the key variables for the investors' decisions and markets' evolution. In addition, there is a significant positive payoff of a more detailed study of such mechanisms inside an extended framework of financial markets as complex systems. In fact, the importance of such study goes beyond the singular problem of the learning processes which are prevailing on financial markets and addresses the critical issue of auto-adaptive mechanisms of these.

References:

[1] Black, F., Noise. *Journal of Finance* 41, 529–544, 1986

[2] Brown, H., Reason, Judgment and Bayes's Law. *Philosophy of Science* 61 (3): 351–369, 1994

[3] Calvet, L., Mandelbrot, B., Fisher, A., A multifractal model of asset returns. *Cowles Foundation Discussion Paper*, 1997, September 15

[4] Cheong, C.W., Mohd Nor, A.H.S., Isa, Z., An Evaluation of Autoregressive Conditional Heteroskedasticity Modeling in Malaysian Stock Market. *WSEAS Transactions on Business and Economics,* Issue 3, Volume 3, ISSN 1109-9526, March 2006

[5] Donaldson, W.H., Hearing before the subcommittee on capital market, insurance and government sponsored enterprises. *Committee on financial services, U.S. House of Representatives*, October 30, 2003

[6] Eisler, Z., Kertész, J., Multifractal model of asset returns with leverage effect. *arXiv:cond-mat/0403767 v2*, 11 May 2004

[7] Feldman, J., Minimization of Boolean complexity in human concept learning. *Nature*, 407:630–633, 2000

[8] Feldman, J., The Simplicity Principle in Human Concept Learning. *Psychology Science* 12: 227–232, 2003

[9] Filis, G., Leon, C., Comovements in the European Financial Markets- Do They Change over Time? A VECM Approach. *WSEAS Transactions on Business and Economics,* Issue 3, Volume 3, ISSN 1109-9526, March 2006

[10] Lo, A.W., The Adaptive Markets Hypothesis, *The Journal of Portfolio Management*, 30th Anniversary Issue 2004, pp. 15-29

[11] Lo, A.W., Reconciling Efficient Markets with Behavioral Finance: The Adaptive Markets Hypothesis, *Social Science Research Network Papers*, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=728864, 2005

[12] Lux, T., The Multi-Fractal Model of Asset Returns: Its Estimation via GMM and Its Use for Volatility Forecasting. *University of Kiel, Working Paper*, 2003

[13] Lux, T., Kaizoji, T. , Forecasting Volatility and Volume in the Tokyo Stock Market: The Advantage of Long Memory Models. *University of Kiel, Working Paper*, 2004

[14] Mandelbrot, B., Hudson, R., The (Mis)behavior of Markets: A Fractal View of Risk, Ruin and Reward. New York: *Basic Books*, 2004

[15] Pan, H.,, Swingtum – A Computational Theory of Fractal Dynamic Swings and Physical Cycles of Stock Market in A Quantum Price-Time Space. *2003 Hawaii International Conference on Statistics and Related Fields*

[16] Ratcliff, R., Comparing Exemplar and Rule-Based Theories of Categorization. *Current Directions in Psychological Science* 15: 9–13, 2006

[17] Rendell, L. A general framework for induction and a study of selective induction. *Machine Learning* 1: 177–226,1986

[18] Sarlija, N., Bensic M., Zekic-Susac, M., Logistic regression, survival analysis and neural networks in modelling customer credit scoring. *WSEAS Transactions on Business and Economics,* Issue 3, Volume 3, ISSN 1109-9526, March 2006