Accepted Manuscript

Investigating market efficiency through a forecasting model based on differential equations

Charlene C. de Resende, Adriano C.M. Pereira, Rodrigo T.N. Cardoso, A.R. Bosco de Magalhães

PII: S0378-4371(17)30066-3
DOI: http://dx.doi.org/10.1016/j.physa.2017.01.057
Reference: PHYSA 17950

To appear in: *Physica A*

Received date: 2 September 2016
Revised date: 20 December 2016

Please cite this article as: C.C. de Resende, A.C.M. Pereira, R.T.N. Cardoso, A.R.B. de Magalhães, Investigating market efficiency through a forecasting model based on differential equations, *Physica A* (2017), http://dx.doi.org/10.1016/j.physa.2017.01.057

*Highlights (for review)

Investigating Market Efficiency through a Forecasting Model based on Differential Equations

Highlights

- A new framework for stock price trend forecast is proposed.
- Information about future price is found in the series analyzed.
- Market efficiency is investigated.
- Application of the model as part of trading strategy is discussed.

Investigating Market Efficiency through a Forecasting Model based on Differential Equations

Charlene C. de Resende[a], Adriano C. M. Pereira[a,c,], Rodrigo T. N. Cardoso[a,b,], A. R. Bosco de Magalhães[a,b,]

[a] *Programa de Pós Graduação em Modelagem Matemática e Computacional- CEFET-MG*
[b] *Departamento de Física e Matemática (DFM) - Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG) - Av. Amazonas 7675 - Nova Gameleira - Belo Horizonte - MG - CEP 30510-000 - Brasil.*
[c] *Departamento de Ciência da Computação (DCC) - Universidade Federal de Minas Gerais (UFMG) - Av. Antônio Carlos - 6627 - Prédio do ICEx - Pampulha - Belo Horizonte - MG - CEP 31270-901 - Brasil.*

Abstract

A new differential equation based model for stock price trend forecast is proposed as a tool to investigate efficiency in an emerging market. Its predictive power showed statistically to be higher than the one of a completely random model, signaling towards the presence of arbitrage opportunities. Conditions for accuracy to be enhanced are investigated, and application of the model as part of a trading strategy is discussed.

Keywords: Stock Markets, Financial Series, Differential Equation Models, Econophysics.

1. Introduction

As it is widely known, according to the efficient market hypothesis (EMH), for rational behavior and market efficiency there is no information in the trading history of an asset that leads to arbitrage opportunities [1]. Nevertheless, a large number of market players are engaged in obtaining such information: in fact, although fundamentalists hold most of stocks, technical traders contribute to the majority of the volume traded [2]. This discussion includes the definition of market efficiency and a non-trivial analysis of market agents rationality; seminal ideas are found in Refs. [3–5]. Fundamental contributions to understanding market dynamics have been developed in the context of behavioral economics and finance, where irrational behavior is studied in detail [6, 7], and in experimental approaches to economics, where the conditions for the market to evolve as predicted by economic theory are investigated [8, 9]. Adaptive market hypothesis [10], which asserts that the market evolves according to natural selection of bounded rational agents, permits return predictability from time to time, and has been argued to furnish satisfactory explanation in some cases.

Email addresses: charlenecassia@gmail.com (Charlene C. de Resende), adrianoc@dcc.ufmg.br (Adriano C. M. Pereira), rodrigoc@des.cefetmg.br (Rodrigo T. N. Cardoso), magalhaes@des.cefetmg.br (A. R. Bosco de Magalhães)

Recent empirical investigations of market efficiency have been carried out in many contexts using a variety of statistical tools (see, for instance [11–26]). Deviations from efficiency were associated to a multiplicity of factors, e.g., operation costs [27], low liquidity [15, 28], information flux [29–31], central bank interventions [32–35], political uncertainties [36] and financial crisis [37]. Knowledge about market inefficiencies are crucial in the context of risk management; indeed, correlations between stocks tend to be greater in periods of market stress, what deteriorate the protection effects of diversification in portfolios in the periods where they are most important [38–41]. On the other hand, noticeable changes in dynamics of financial markets may be used by governments and market players to detect early signs of financial crises [42, 43]. Understanding the way the market deviates from the standard models can also be employed to build profitable trading strategies [44–46].

There are in the literature a plenty of articles focused on market dynamics forecast (a survey of papers based on conventional methods is found in Ref. [47] and on soft computing methods in Ref. [48]). Forecast models tend to be improved when collective behavior is taken into account [49]. Co-movement of asset prices is studied in various ways [38–43, 49–53]. Although forecasts about trend reversals of asset prices at a given time using information about other assets at the same time can achieve high accuracies, such accuracies drop drastically for forecasts on future time [49]; this is in line with EMH. Predicting volatility also does not contradict EMH: in fact, we can find methods based on historical series or on option prices with satisfactory results (see Ref. [54] for a review, recent examples are in refs. [55–57]). On the other hand, predicting the sign of asset returns may be related to market inefficiency. Efforts towards predicting return signs are found in refs. [58–61]. In Ref. [62], the relation between forecastability of volatility and of return sing is explored. Recently, the behavior of internet users has been used as a probe in the market. Significant correlations between transaction volumes and search volumes in the search engine Google were found [63]. Behavioral data from the internet was also used as a key ingredient in trading strategies [45, 46, 64]. In the same context, neural networks, fuzzy logic and support vector machines are popular tools which are often used in conjunction (e.g., refs. [65–69]). In the present contribution, we propose a new tool for price trend forecast which can be employed as part of trading strategies and to investigate market efficiency. It is based on differential equations and take advantage of collective behavior by using data about multiple stocks.

Our main goal here is to study market efficiency by using the proposed model. We show that its success rate in price trend forecasting is consistently higher than the one of a complete random model. This makes explicit the existence of information about future prices in the series analyzed, what is empirical evidence of market inefficiency. Our model is founded on the idea that the dynamics of the deviation of the price of an asset from its fair price depends on such deviation and also on the respective deviations of other prices. The dependence is modeled through a system of linear differential equations. In view of simplicity, linear relations are natural choice for a first exploration of such a differential equation based model. However, we do not expect that simple linear relations can exactly describe the dynamics of stock prices: the model proposed here was employed as an approximation capable of providing *some* information about future movements. The inclusion of nonlinear terms and the investigation of their effect in the model accuracy is subject matter for future research. We chose to work in a scenario were predictability is more probable to be observed: since inefficiencies tend to be stronger in emerging markets [13, 14], we tested our model with stocks from the Brazilian market BM&FBOVESPA. We also used intraday data, where a richer panorama of correlations is expected [70].

Two versions of the model are described in Section 2: one of them uses a minimum set of data points to calculate the parameters of the model at each time step; the other one is fitted over larger intervals. While calculation of the former is relatively direct, involving the inversion of a matrix as the most computationally expensive step, the latter demanded the utilization of nonlinear optimization techniques. Results of different ways of applying the model to empirical data are displayed in Section 3, where they are statistically compared to outcomes of completely random forecasts. In the same section, we calculate some evaluation metrics for the forecasts of the model in a version designed for application in trading strategies. Final remarks are found in Section 4.

2. Model

2.1. Exact Fit Model

Let us consider N assets and denote the price of each one at time t by $x_k(t)$ $(k = 1, 2, \cdots, N)$. We work here with closing prices of candles corresponding to constant time intervals with size Δt, coming from empirical data. Aiming at setting up the differential equations based model, we define t as a continuous variable whose sampling is discrete. Deviation from the fair value is defined as

$$X_k(t) = \ln x_k(t) - \ln \bar{x}_k(t). \tag{1}$$

Of course, the fair value $\bar{x}_k(t)$ is an idealized quantity; here, we use a moving average as an approximation for it:

$$\bar{x}_k(t) = \sum_{\tau=t-S}^{t-1} x_k(\tau). \tag{2}$$

This model is built as a system of linear differential equations:

$$\frac{d}{dt}\mathbf{X}(t) = \mathbf{A}\mathbf{X}(t), \tag{3}$$

where

$$\mathbf{X}(t) = \begin{pmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ X_N(t) \end{pmatrix}, \qquad \mathbf{A} = \begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,N} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ A_{N,1} & A_{N,2} & \cdots & A_{N,N} \end{pmatrix}. \tag{4}$$

If $\mathbf{A}$ does not depend on time, the solutions can be written in the form

$$\mathbf{X}(t_0 + t) = e^{t\mathbf{A}}\mathbf{X}(t_0). \tag{5}$$

In order to build the prediction model, we calculate matrices $\mathbf{A}$ from deviations X_k. Indeed, for computation of each matrix $\mathbf{A}$, we must consider Eq. (5) with N different pairs (t_0, t) that lead to N independent equations. This can be achieved with the help of the square matrices

$$\mathbf{X_N}(t) = \begin{pmatrix} X_1(t-N+1) & \cdots & X_1(t-1) & X_1(t) \\ X_2(t-N+1) & \cdots & X_2(t-1) & X_2(t) \\ \vdots & \ddots & \vdots & \vdots \\ X_N(t-N+1) & \cdots & X_N(t-1) & X_N(t) \end{pmatrix}, \tag{6}$$

which contain values of deviations spaced by one unity of time and permit us to write

$$\mathbf{X_N}(t) = e^{\mathbf{A}}\mathbf{X_N}(t-1), \tag{7}$$

what lead us to

$$\mathbf{A} = \ln\left[\mathbf{X_N}(t)\mathbf{X_N}^{-1}(t-1)\right]. \tag{8}$$

Let us stress that matrix $\mathbf{A}$ is assumed to be constant in the calculation above, and defines a linear system whose solution fall in the points given in $\mathbf{X_N}(t)$ and $\mathbf{X_N}(t-1)$ corresponding to $N+1$ instants of time. Let us also note that $\mathbf{A}$ can vary with the choice of t in Eq. (8). Thus, we use the notation $\mathbf{A}(t)$ hereafter to denote the matrix $\mathbf{A}$ calculated in the time window that ends in t.

Each matrix $\mathbf{A}(t)$ is employed in the attempt to get information about the deviations concerning the next step by means of equation

$$\hat{\mathbf{X}}_\mathbf{N}(t+1) = e^{\mathbf{A}(t)}\mathbf{X_N}(t). \tag{9}$$

Forecasted deviations are found in the N^{th} column of $\hat{\mathbf{X}}_{\mathbf{N}}(t+1)$. If $\mathbf{A}(t) = \mathbf{A}(t+1)$, then $\hat{\mathbf{X}}_{\mathbf{N}}(t+1) = \mathbf{X}_{\mathbf{N}}(t+1)$. In this case, the real deviations $X_k(t+1)$ could be found in matrix $\hat{\mathbf{X}}_{\mathbf{N}}(t+1)$. However, $\mathbf{A}(t)$ varies with time, therefore we do not expect very accurate predictions (this would imply unrealistic arbitrage opportunities). What we investigate here is whether *some* information about deviations $X_k(t+1)$ can be obtained by using Eq. (9). Specifically, we compare price trend forecasts based on Eq. (9) with completely random predictions.

2.2. Approximate Fit Model

In the model presented above, if $\mathbf{X}_{\mathbf{N}}(t-1)$ has an inverse we can find matrices $\mathbf{A}(t)$ perfectly fitted to Eq. (7), i.e., the system defined this way presents solutions that fall in all points in the interval $[t-N, t]$. Now, we change the model seeking for matrices $\mathbf{B}(t)$ that are approximately fitted to larger intervals. This procedure leads to matrices with lesser variation at each time step. As we will see, the predictive power of the model increases with such a variation decrease.

Let us start by recalling the usual distance between two $N \times N$ matrices $\mathbf{C}$ and $\mathbf{D}$:

$$d(\mathbf{C}, \mathbf{D}) = \sqrt{\sum_{i=1}^{N} \sum_{j=1}^{N} \left| [\mathbf{C}]_{i,j} - [\mathbf{D}]_{i,j} \right|^2}; \tag{10}$$

here, we adopt the notation according to which $[\mathbf{C}]_{i,j}$ ($[\mathbf{D}]_{i,j}$) is the element of the matrix $\mathbf{C}$ ($\mathbf{D}$) in the i^{th} line and j^{th} column. Considering a calculation analogous to the one in Eq. (7), but maintaining the same matrix in the exponent for different intervals, we are lead, for each value of t, to

$$\begin{aligned} \mathbf{Y}_{\mathbf{N}}(t) &= e^{\mathbf{B}(t)} \mathbf{X}_{\mathbf{N}}(t-1), \\ \mathbf{Y}_{\mathbf{N}}(t-1) &= e^{\mathbf{B}(t)} \mathbf{X}_{\mathbf{N}}(t-2), \\ &\vdots \\ \mathbf{Y}_{\mathbf{N}}(t-R+1) &= e^{\mathbf{B}(t)} \mathbf{X}_{\mathbf{N}}(t-R), \end{aligned} \tag{11}$$

with integer R. We search for the matrix $\mathbf{B}(t)$ that leads to $\mathbf{Y}_{\mathbf{N}}(\tau)$ as close as possible to $\mathbf{X}_{\mathbf{N}}(\tau)$ for τ varying from $t-R+1$ to t, namely, the matrix $\mathbf{B}(t)$ that minimize the sum of the squares of the distances between the matrices $\mathbf{Y}_{\mathbf{N}}(\tau)$ and $\mathbf{X}_{\mathbf{N}}(\tau)$ $(t-R+1 < \tau < t)$. Thus, our objective function to be minimized is

$$f(\mathbf{B}(t)) = \sum_{\tau=(t-R+1)}^{t} \sum_{i=1}^{N} \sum_{j=1}^{N} \left| [\mathbf{X}_{\mathbf{N}}(\tau)]_{i,j} - \left[e^{\mathbf{B}(t)} \mathbf{X}_{\mathbf{N}}(\tau-1) \right]_{i,j} \right|^2. \tag{12}$$

As the optimization tool we use a hybrid method consisting in the combination of gradient and conjugate gradient methods [71], with the step size given by golden section method and stopping criterion based on error stabilization (variation less than 10^{-9}). The average of the matrices $\mathbf{A}$ relative to the interval $[t-R+1, t]$,

$$\mathbf{B}_0(t) = \frac{1}{R}\sum_{\tau=t-R+1}^{t} \mathbf{A}(\tau), \tag{13}$$

is employed as the initial value in the optimization process.

The exact model can be assumed as a particular case, corresponding to $R = 1$, of the model described in this section. Therefore, we will not distinguish matrices $\mathbf{A}$ from matrices $\mathbf{B}$ in what follows. Equation

$$\hat{\mathbf{X}}_{\mathbf{N}}(t+1) = e^{\mathbf{B}(t)}\mathbf{X}_{\mathbf{N}}(t) \tag{14}$$

generalizes the forecast model given in Eq. (9).

3. Application of Forecast Models

Deviations from efficiency are more likely to be found in emerging markets than in developed ones [13, 14]. Aiming to test our forecast model in an emerging market, we chose three stocks from the Brazilian market BM&FBOVESPA: Itaúsa (ITSA4), Bradesco (BBDC4) and Usiminas (USIM5). As will be clear below, this choice also permit us to investigate the importance of linear correlation between stocks for the predictive power of the model. The data we work with correspond to adjusted closing prices of 5 minutes candles collected in an interval starting in 03/10/2014 and ending in 07/01/2014. From this set, we took off the points of the days when the market did not operate full time. The fair prices $\bar{x}_k(t)$ were computed as averages taken in the hour just before the instant t (namely, $S = 12$ in Eq. (2)). In order to avoid the influence of previous day in the calculation of the deviations $X_k(t)$, we discarded the deviations relative to the first negotiation hour of each day. With the points that lasted, we computed $4,500$ matrices $\mathbf{B}(t)$ for each model employed, and also the Pearson correlation coefficient between ITSA4 and BBDC4 (0.3804), ITSA4 and USIM5 (0.0894), and BBDC4 and USIM5 (0.0689). Notice that the linear correlation between ITSA4 and BBDC4 is far more relevant than the other. As we shall see, the predictive power of the model is higher for more correlated assets.

The model proposed here was analyzed with respect to its ability to forecast, at time t, if the price at $t + 1$ will be higher or lower than at t, i.e., we focused only on the trend and not on the value of the prices forecasted. To this end, we first used Eq. (14) to compute predicted deviations. The forecasted prices were calculated from the deviations; the forecasted trends were computed from the prices. Our primary objective is to verify the presence of information in the temporal series of prices. Thus, we compared the accuracy of our forecasts with the accuracy of random predictions that the price will go up or down (with probability 0.5

each one), i.e., forecasts similar to those of a fair coin. Since the fair coin model never predicts stability for adjacent prices, it has no chance of providing a correct forecast in those cases; indeed, if we do not consider computational limits, this chance is also null for our differential equation based model, given its continuous nature. If we remove, from the set U of $4,500$ trend forecasts for a given stock, those concerning the intervals where the actual variation of the price was zero, we get the subset $\bar{U}$ for which the fair coin model gives, in average, the right trend prediction half of the times. Deviations from this proportion, observed for the differential equation based forecast model, indicate the presence of information about future returns in the time series. In order to investigate if the size of a predicted return carries information about the trend, we divided U in four subsets U_1, U_2, U_3 and U_4 with $1,125$ elements each one, in such a way that any return regarding a trend forecast in U_i is lower, in modulus, than the ones connected to the forecasts in U_{i+1} (i = 1, 2 or 3). Thus, for each asset, U_1 contains a quarter of the trend forecasts, the ones associated with the lowest return moduli and U_4, the fourth part of the trend forecasts connected to the highest return moduli; forecasts related to intermediate return absolute values are included in the sets U_2 and U_3. By removing from U_1, U_2, U_3 and U_4 the forecasts associated to the intervals in which the actual variation is null, we get the sets $\bar{U}_1$, $\bar{U}_2$, $\bar{U}_3$ and $\bar{U}_4$ where the differential equation model can be compared to the fair coin model.

In tables (1, 2 and 3), we show the accuracy

$$a = \frac{n_c}{n}, \tag{15}$$

where n_c is the number of correct trend forecasts and n is the total number of forecasts, of the trend forecasts of the differential equation based model, for the three stocks studied. Such accuracies were calculated for the sets $\bar{U}$, $\bar{U}_1$, $\bar{U}_2$, $\bar{U}_3$ and $\bar{U}_4$, and for different sizes of the intervals used to find the matrices $\mathbf{B}(t)$ (R varying from 1 to 10). The hypothesis tests that the accuracies are statistically different from 0.5 were made:

$$\begin{cases} H_0 : a = 0.5 \\ H_1 : a \neq 0.5 \end{cases}.$$

For each case in tables (1, 2 and 3), a z-test was performed in which the stochastic variable is the sequence of successes and fails of the model trend forecasts. The assumptions of normality and homoscedasticity are verified by the sizes of the samples [72]; the assumption of independence was tested with run test for randomness. Tables (1, 2 and 3) display the p-values of z-tests. The p-values less than or equal to 0.05 are bold, corresponding to the cases in which our model wins the fair coin model, and the colored cells correspond to those with p-values less than or equal to 0.05 in run test, meaning lack of independence. Notice that in both cases the sequences of successes and fails of the model forecasts are statistically different than the ones from a fair coin. In 41% of the cases in tables (1, 2 and 3) the sample is considered to be random (not colored cell) and the accuracy is statistically equal to 0.5 (not bold numbers), which represents the case of a fair coin. However, if we take into account only the correlated stocks ITSA4 and BBDC4, this percentage drops to 32% and 34%, respectively, and, if we compute exclusively in the sets $\bar{U}_4$, we find a single *fair coin* case for BBDC4 and none for ITSA4. For USIM5, three *fair coin* cases are found in $\bar{U}_4$, all of them for $R < 6$. Generally speaking, the model results tend to move away from the ones of a fair coin for the correlated stocks ITSA4 and BBDC4 and when R increases; this is more evident for the sets $\bar{U}_3$ and $\bar{U}_4$. The p-values found for those cases indicate that the probability of the fair coin model to generate equivalent results is very low. On the other hand, results concerning USIM5 are less significant.

The gain obtained employing the approximate fit model may be related to the decrease, with the increase of R, of the variation of the matrices calculated at each step. The computation of each matrix $\mathbf{B}(t)$ is performed using points distributed at $N + R$ instants of time. Thus, increased R leads to larger amount of data shared in the calculation of two matrices $\mathbf{B}$ regarding adjacent times. In Figure 1, we show the distributions of the distances $d(\mathbf{B}(t-1), \mathbf{B}(t))$ for different values of R: notice that the increase of R actually leads, in average, to lesser distances. Great variability of the matrices deteriorate the model predictive power. For the range of values of R studied, as we calculate matrices that fit over longer intervals, the variability decreases and the information about the future price becomes more relevant.

Table 1: Statistics of forecasts for ITSA4 in the sets $\bar{U}$, $\bar{U}_1$, $\bar{U}_2$, $\bar{U}_3$ and $\bar{U}_4$ for different values of the parameter R.

ITSA4	R	1	2	3	4	5	6	7	8	9	10
$\bar{U}$	n	3291	3291	3291	3291	3291	3291	3291	3291	3291	3291
	a	0.5090	0.5332	0.5366	0.5369	0.5360	0.5426	0.5420	0.5378	0.5454	0.5448
	p	0.3017	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**
$\bar{U}_1$	n	803	784	772	769	789	775	774	782	785	789
	a	0.4770	0.5089	0.5168	0.5058	0.5183	0.5277	0.5284	0.5063	0.5235	0.5031
	p	0.1923	0.6242	0.3524	0.7490	0.3078	0.1236	0.1142	0.7264	0.1902	0.8650
$\bar{U}_2$	n	812	808	805	821	809	811	813	794	809	827
	a	0.5170	0.5074	0.5366	0.5286	0.5055	0.5092	0.5141	0.5314	0.5302	0.5296
	p	0.3252	0.6744	**0.0384**	0.1032	0.7566	0.6030	0.4238	0.0780	0.0870	0.0890
$\bar{U}_3$	n	809	848	845	827	828	820	815	827	816	805
	a	0.5093	0.5731	0.5408	0.5417	0.5471	0.5597	0.5607	0.5368	0.5551	0.5540
	p	0.5967	**<0.001**	**0.0178**	**0.0168**	**0.0068**	**<0.001**	**<0.001**	**0.0348**	**<0.001**	**0.0022**
$\bar{U}_4$	n	867	851	869	873	865	885	889	888	881	870
	a	0.5306	0.5404	0.5500	0.5681	0.5699	0.5706	0.5624	0.5720	0.5698	0.5885
	p	0.0715	**0.0188**	**0.0032**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**

Note: The symbol n represents the total number of forecasts, a is the accuracy and p is the p-value for hypothesis test assuming as the null hypothesis that the forecast model gives the right answer half of the times in the sets analyzed. Highlighted in bold 28 p-values less than 0.05, corresponding to nearly 56% of the cases. The colored cells correspond to the cases where the sample violates the premise of independence of data (30%).

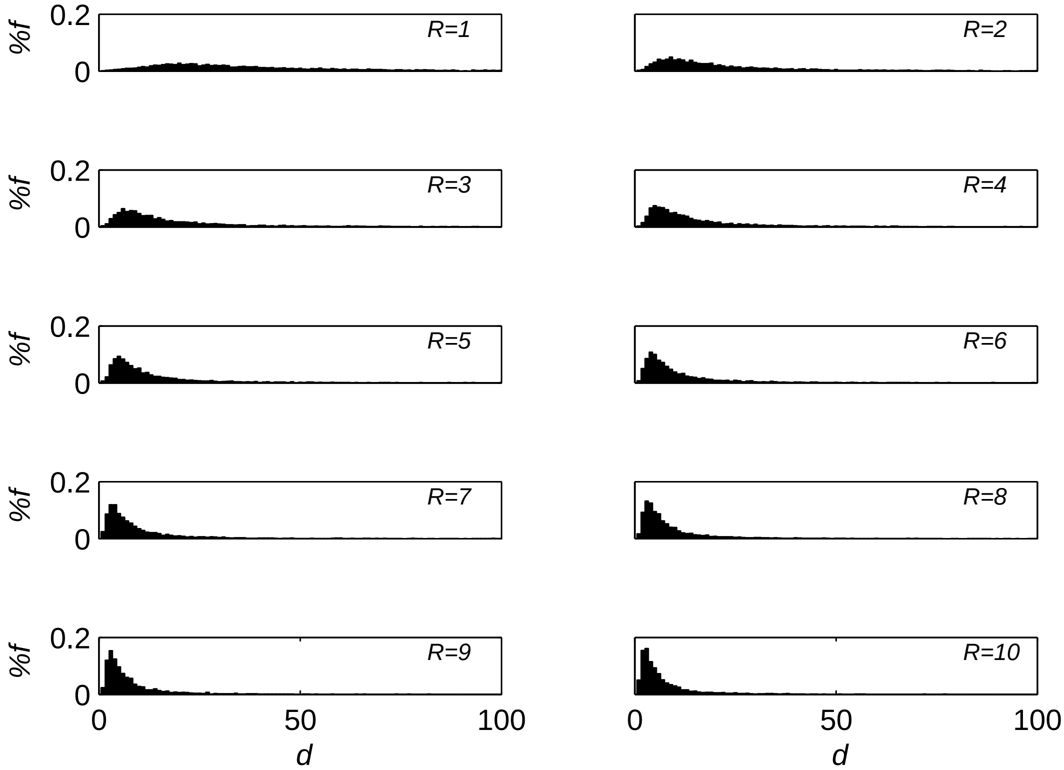

Figure 1: Distributions of distances between matrices **B** regarding adjacent times for different values of the parameter R. Higher values of R correspond to larger amounts of shared data in the calculation of adjacent matrices, which leads to smaller distances between them.

Table 2: Statistics of forecasts for BBDC4 in the sets $\bar{U}$, $\bar{U}_1$, $\bar{U}_2$, $\bar{U}_3$ and $\bar{U}_4$ for different values of the parameter R.

BBDC4	R	1	2	3	4	5	6	7	8	9	10
$\bar{U}$	n	4118	4118	4118	4118	4118	4118	4118	4118	4118	4118
	a	0.5078	0.5128	0.5223	0.5269	0.5356	0.5337	0.5359	0.5378	0.5356	0.5371
	p	0.3167	0.1001	**0.0042**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**
$\bar{U}_1$	n	1012	1014	1013	1002	1016	996	1007	1011	1011	1011
	a	0.4881	0.5000	0.5123	0.4940	0.4960	0.5050	0.5004	0.5222	0.5103	0.5133
	p	0.3764	1.000	0.4354	0.7114	0.8026	0.7566	0.9840	0.1586	0.5156	0.4010
$\bar{U}_2$	n	1029	1026	1024	1023	1005	1023	1015	1018	1013	1016
	a	0.4791	0.5136	0.4931	0.5083	0.5255	0.5161	0.5172	0.5147	0.5074	0.5068
	p	0.1799	0.3844	0.6600	0.5962	0.1096	0.3078	0.2758	0.3524	0.6384	0.6672
$\bar{U}_3$	n	1034	1041	1045	1038	1043	1040	1040	1037	1042	1034
	a	0.5213	0.5120	0.5244	0.5385	0.5465	0.5326	0.5298	0.5313	0.5566	0.5570
	p	0.1707	0.4412	0.1164	**0.0132**	**0.0028**	**0.0358**	0.0548	**0.0444**	**<0.001**	**<0.001**
$\bar{U}_4$	n	1043	1037	1036	1054	1054	1059	1056	1052	1052	1057
	a	0.5417	0.5255	0.5588	0.5654	0.5730	0.5588	0.5937	0.5817	0.5665	0.5695
	p	**0.0070**	0.1010	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**

Note: The symbol n represents the total number of forecasts, a is the accuracy and p is the p-value for hypothesis test assuming as the null hypothesis that the forecast model gives the right answer half of the times in the sets analyzed. Highlighted in bold 23 p-values less than 0.05, corresponding to nearly 46% of the cases. The colored cells correspond to the cases where the sample violates the premise of independence of data (24%).

Table 3: Statistics of forecasts for USIM5 in the sets $\bar{U}$, $\bar{U}_1$, $\bar{U}_2$, $\bar{U}_3$ and $\bar{U}_4$ for different values of the parameter R.

USIM5	R	1	2	3	4	5	6	7	8	9	10
$\bar{U}$	n	3450	3450	3450	3450	3450	3450	3450	3450	3450	3450
	a	0.5064	0.5095	0.5081	0.5057	0.5098	0.5139	0.5133	0.5127	0.5153	0.5200
	p	0.4521	0.2670	0.3422	0.5092	0.1032	0.1032	0.1188	0.1362	0.0734	**0.0192**
$\bar{U}_1$	n	811	829	817	823	821	817	836	846	845	843
	a	0.5080	0.5211	0.4895	0.5066	0.4993	0.5030	0.4940	0.5023	0.5053	0.5077
	p	0.6486	0.2262	0.5486	0.7114	0.9680	0.8650	0.7338	0.8966	0.7642	0.6600
$\bar{U}_2$	n	833	838	834	844	838	855	831	825	840	850
	a	0.4862	0.4701	0.5191	0.5000	0.5000	0.5087	0.5162	0.5115	0.5011	0.5200
	p	0.4256	0.0836	0.2714	1	1	0.6170	0.3524	0.5092	0.9522	0.2460
$\bar{U}_3$	n	872	874	882	863	876	860	870	868	846	838
	a	0.4966	0.5114	0.5056	0.5005	0.5159	0.5058	0.5000	0.4907	0.5153	0.5190
	p	0.8408	0.5028	0.7414	0.9840	0.3472	0.7338	1	0.5892	0.3734	0.2714
$\bar{U}_4$	n	934	909	917	919	915	918	913	911	919	919
	a	0.5321	0.5333	0.5169	0.5157	0.5224	0.5359	0.5410	0.5444	0.5375	0.5321
	p	**0.0497**	**0.0456**	0.3078	0.3422	0.1770	**0.0300**	**0.0136**	**0.007**	**0.0226**	0.0524

Note: The symbol n represents the total number of forecasts, a is the accuracy and p is the p-value for hypothesis test assuming as the null hypothesis that the forecast model gives the right answer half of the times in the sets analyzed. Highlighted in bold 7 p-values less than 0.05, corresponding to nearly 14% of the cases. The colored cells correspond to the cases where the sample violates the premise of independence of data (32%).

According to tables (1, 2 and 3), the probability of obtaining correct trend forecasts tends to enhance when the absolute value of the associated predicted return grows, since the model predictive power is generally higher for the forecasts belonging to the sets $\bar{U}_3$ and $\bar{U}_4$. In order to better analyze this, we have built figures 2, 3 and 4, where each accuracy is computed in a set $\bar{U}_{k,\Delta}$ of trend forecasts constructed as follows: the trend forecasts are ordered according to the absolute value of the predicted returns associated to them (from smallest to largest); $\bar{U}_{k,\Delta}$ is the set of forecasts from the position k to the position $k + \Delta - 1$, where Δ is the number of elements of $\bar{U}$ divided by four (rounded to the closest integer). In other words, $\bar{U}_{k,\Delta}$ corresponds to a sliding window of size Δ which encompasses trend forecasts associated to growing predicted returns as k increases. Although the curves are not monotonic, clearly we observe a tendency that the accuracy grows with increasing k, which is confirmed in Table 4 showing the angular coefficients β resulting from linear regressions performed on the points in figures 2, 3 and 4: notice that all coefficients are positive;

their magnitude orders (10^{-5} or 10^{-6}) are related to the typical range of variation of the accuracies observed here (10^{-1}) and the number of points in figures 2, 3 and 4, which is of the order of 10^3. The coefficient of determination r^2 of each adjust is also displayed, showing that the linear approximations tend to perform better when the model performs better (ITSA4 and BBDC4).

A procedure analogous to that described above can be used to investigate the influence of volatility on forecasts accuracy. For each prediction from the 101st of each set, we computed the volatility in the interval of 100 points just before it. Then we sorted the forecasts according to their associated volatility in ascending order and computed the accuracy in sliding windows similar to $\bar{U}_{k,\Delta}$. Here, the window size Δ for each stock is the number of forecasts with associated volatility divided by four (rounded to the closest integer) and k starts at 101. The dependence of accuracy on k for $R = 10$ is shown in Fig. 5; no qualitative differences were observed for other values of R. Angular coefficients β and coefficients of determination r^2 resulting from linear regressions performed on the points of Fig. 5 and on the corresponding points for other values of R are found in Table 5. Given the sign variation of those coefficients, we conclude that our computation of volatility did not furnish a clue for accuracy improvement.

Let us now focus on the use of the model as part of a trade strategy. According to the discussion above, such a strategy should employ the trend forecasts related to the predicted returns with the higher absolute values, which tend to be more accurate. In tables (1, 2 and 3), this corresponds to the sets $\bar{U}_4$. For application in trade strategy, one has to be able to classify a predicted return as part of the set with the higher absolute values using only past data. Notice that the classification of a forecast in the set $\bar{U}_4$ is performed employing all the series, i.e., for a point in the middle of the series, future data are used to do so. Bearing this in mind, we built the sets $\bar{V}_4$, analogous to $\bar{U}_4$, with the difference that the classification of each trend forecast as belonging or not to them is performed by considering only the 100 forecasts immediately preceding it, instead of the whole series. Statistics for the sets $\bar{V}_4$ are presented in Table 6: as they are not qualitatively different than the statistics of $\bar{U}_4$, we conclude that the use of future data to classify forecasts in $\bar{U}_4$ is not the source of the gain in accuracy observed for these sets.

Still thinking in using the model in trading strategies, we built new sets of trend forecasts obtained from a voting method involving, at every step, only the trend forecasts of the sets $\bar{V}_4$ concerning the top half of R values (R ranging from 6 to 10). This choice of voters is natural, since the accuracies tend to be greater for higher values of R, as well as for predictions in the set $\bar{V}_4$. For the steps where there is a draw or no voters, no forecast is generated. Confusion matrices for these new sets are shown in Table 7, and evaluation metrics derived from them, in Table 8. The accuracies reached are close to the accuracies of the predictions

Table 4: Slopes of the lines of best fit (β) of the graphs in figures 2, 3 and 4 and corresponding coefficients of determination (r^2).

	R	**1**	**2**	**3**	**4**	**5**
ITSA4	β	1.88×10^{-5}	2.89×10^{-5}	9.05×10^{-6}	1.58×10^{-5}	3.34×10^{-5}
	r^2	0.7936	0.6579	0.6084	0.6170	0.7915
BBDC4	β	2.02×10^{-5}	3.10×10^{-6}	1.90×10^{-5}	2.87×10^{-5}	2.19×10^{-5}
	r^2	0.7015	0.0943	0.6477	0.8610	0.7832
USIM5	β	1.24×10^{-5}	2.50×10^{-5}	5.83×10^{-6}	7.37×10^{-6}	1.52×10^{-5}
	r^2	0.4747	0.4000	0.1692	0.0953	0.7180

	R	**6**	**7**	**8**	**9**	**10**
ITSA4	β	3.15×10^{-5}	1.75×10^{-5}	2.27×10^{-5}	2.13×10^{-5}	2.73×10^{-5}
	r^2	0.7197	0.5712	0.6620	0.8415	0.8850
BBDC4	β	1.98×10^{-5}	2.54×10^{-5}	2.16×10^{-5}	2.40×10^{-5}	1.90×10^{-5}
	r^2	0.6933	0.7544	0.5203	0.8625	0.6197
USIM5	β	8.16×10^{-6}	1.28×10^{-5}	5.09×10^{-6}	8.39×10^{-6}	4.64×10^{-6}
	r^2	0.3281	0.5542	0.0997	0.3581	0.2206

Note: All coefficients β are positive. Their magnitude orders (10^{-5} or 10^{-6}) are related to the typical range of variation of the accuracies (10^{-1}) and the number of points (of the order of 10^3) in figures 2, 3 and 4.

Table 5: Slopes of the lines of best fit (β) and coefficients of determination (r^2) of the graphs in figure 5 and the corresponding graphs for other values of R.

	R	1	2	3	4	5
ITSA4	β	4.29×10^{-6}	-2.28×10^{-5}	-3.94×10^{-6}	-8.74×10^{-6}	-2.02×10^{-5}
	r^2	0.1618	0.7513	0.0437	0.2453	0.5032
BBDC4	β	6.50×10^{-6}	-5.86×10^{-6}	6.36×10^{-6}	-2.22×10^{-6}	4.95×10^{-6}
	r^2	$0,3134$	$0,3166$	0.2865	0.0220	0.0839
USIM5	β	-1.35×10^{-5}	6.05×10^{-6}	1.58×10^{-5}	2.43×10^{-5}	1.39×10^{-5}
	r^2	0.2796	0.1887	0.6849	0.8808	0.6235

	R	6	7	8	9	10
ITSA4	β	-2.24×10^{-5}	-2.91×10^{-5}	-1.50×10^{-5}	-2.71×10^{-5}	-1.35×10^{-5}
	r^2	0.6921	0.6724	0.3844	0.6755	0.5594
BBDC4	β	1.88×10^{-5}	1.20×10^{-5}	-5.20×10^{-6}	4.83×10^{-6}	-2.89×10^{-6}
	r^2	0.6902	0.4065	0.2980	0.0638	0.0463
USIM5	β	1.71×10^{-5}	2.04×10^{-5}	4.08×10^{-6}	1.69×10^{-6}	1.45×10^{-5}
	r^2	0.7107	0.6014	0.0241	0.0044	0.4707

Note: The variation in the sign of the coefficients β indicate that the increase in volatility is not directly related to improvement in model accuracy.

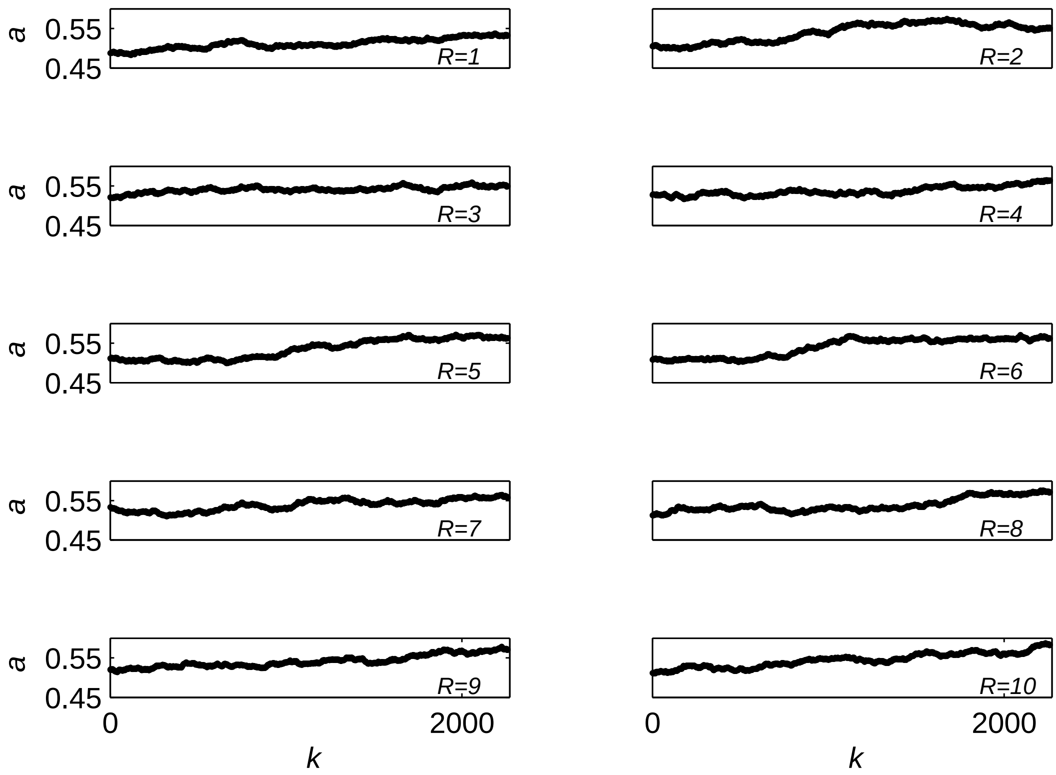

Figure 2: Accuracy of the forecasts in the set $\bar{U}_{k,\Delta}$ as a function of k for ITSA4 (Δ = 823). Results are shown for R varying from 1 to 10. Although the curves above are not monotonic, they show the tendency that the accuracy grows with increasing k. This is confirmed by the positive angular coefficients resulting from linear regressions displayed in Table 4.

that took part in the voting; the advantage achieved with the voting is the number of predictions obtained, which is greater than in any set $\bar{U}_4$ or $\bar{V}_4$. Sensitivity and specificity are related to the predictive power of the model in the subsets of *actual* positive and negative returns, respectively. As we see in Table 8, sensitivity is higher than specificity for ITSA4, and lower for BBDC4 and USIM5; thus, we can not say the model fits better in one subset than in the other. Negative and positive predictive values give the success rate of forecasts for *predicted* positive and negative returns, respectively. Here again we can not assign to the model a better performance in any subset. The results for F1 scoreA (F1 scoreB), which correspond to the harmonic mean of sensitivity and positive predictive value (specificity and negative predictive value), also show the similarity of the model performance for all classes. Notice that the fraction of positive and negative actual

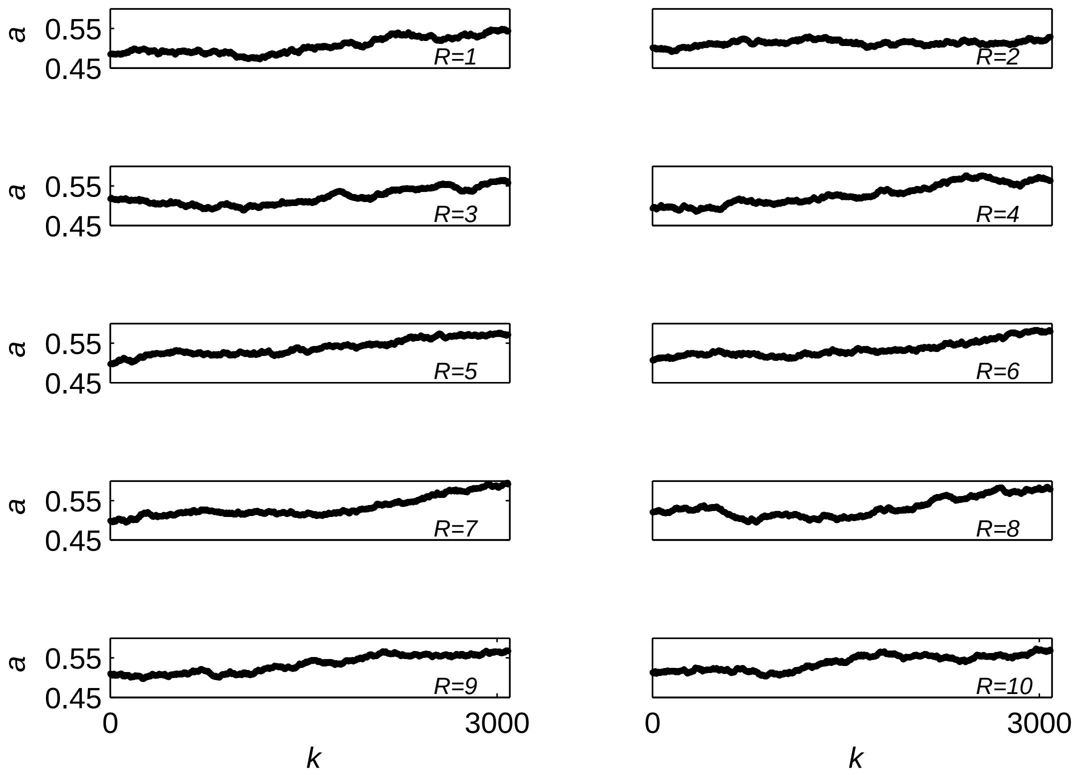

Figure 3: Accuracy of the forecasts in the set $\bar{U}_{k,\Delta}$ as a function of k for BBDC4 ($\Delta = 1030$). Results are shown for R varying from 1 to 10. Although the curves above are not monotonic, they show the tendency that the accuracy grows with increasing k. This is confirmed by the positive angular coefficients resulting from linear regressions displayed in Table 4.

returns in each series analyzed is close to 0.5.

As a last investigation, we focus on the influence of the variation of the number of assets N on accuracy. For this analysis, we choose the five stocks that integrate Ibovespa's theoretical portfolio (the most important indicator in Brazilian stock market) with the highest participation in the first four-month period of 2014: Itaú Unibanco (ITUB4), Petrobras (PETR4), Vale (VALE5), Bradesco (BBDC4) and Ambev (ABEV3), in descending order of participation. The data corresponds to the same period and sampling procedure of the previous investigation. Forecasts were performed according to the same scheme for the preparation of Table 8, i.e., by means of votings involving the predictions in the sets of the type $\bar{V}_4$ concerning R ranging from 6 to 10. Accuracies are displayed in Table 9. Relevant gain in accuracy is observed when N goes from 2 to 3,

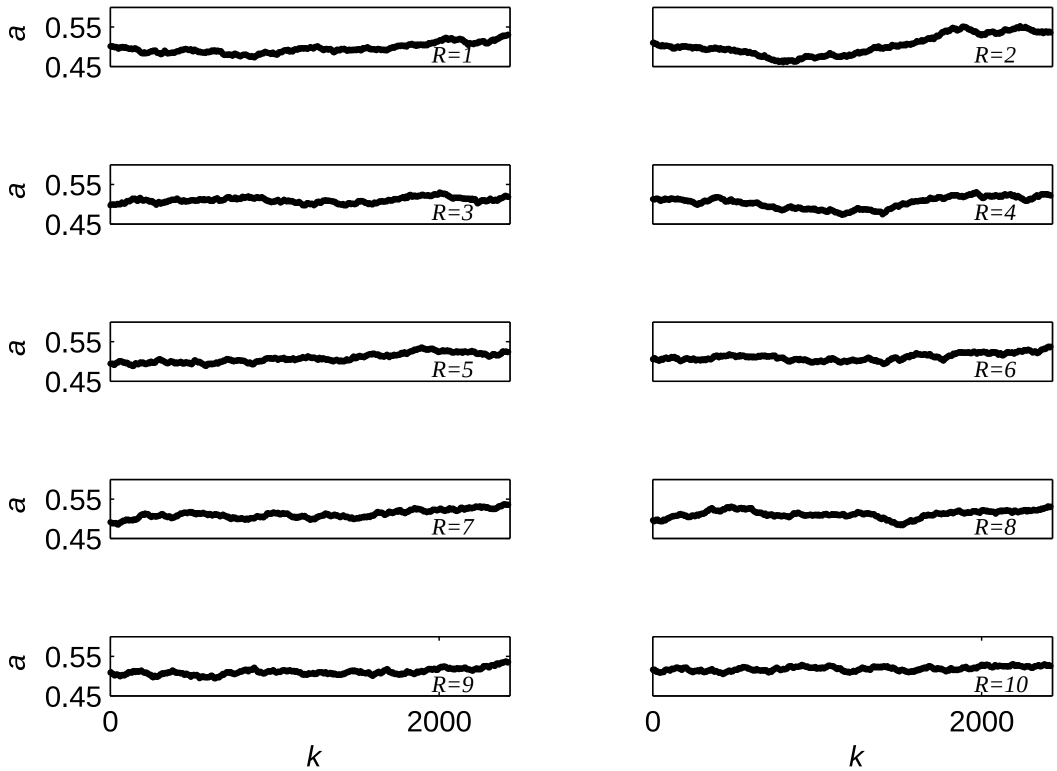

Figure 4: Accuracy of the forecasts in the set $\bar{U}_{k,\Delta}$ as a function of k for USIM5 (Δ = 863). Results are shown for R varying from 1 to 10. Although the curves above are not monotonic, they show the tendency that the accuracy grows with increasing k. This is confirmed by the positive angular coefficients resulting from linear regressions displayed in Table 4.

but no consistent improvement is noted for further increase in such number.

4. Final Remarks

We proposed a model for stock price trend forecast based on differential equations that was used as a tool for the investigation of market efficiency. For a realistic situation, it is not expected that any model can accurately predict future prices. In the present case, this is reflected in the variability of the matrices that define the differential equations used to perform the forecasts at every time step. We expect, however, that even with such variation, some predictive power remains. Although predictability in trend forecasting does not always lead to arbitrage opportunity, which depends on factors such as the distributions of returns,

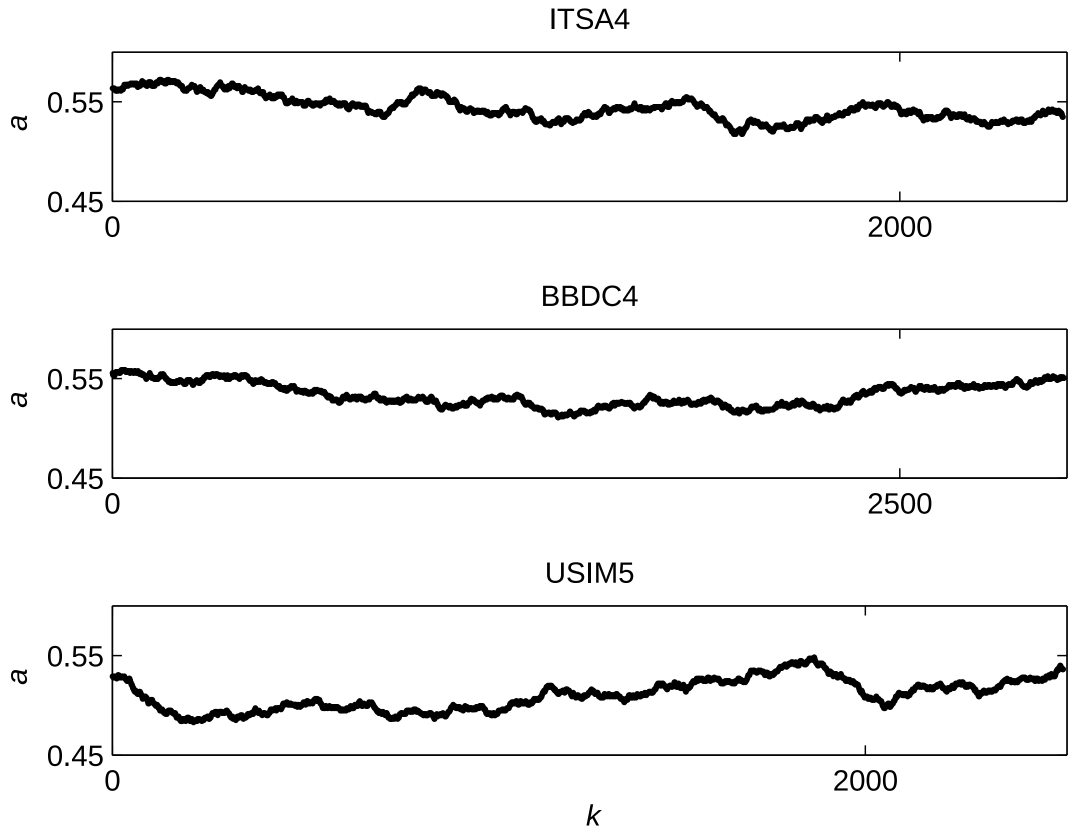

Figure 5: Accuracy of the forecasts in the set $\bar{U}_{k,\Delta}$ as a function of k for ITSA4 ($\Delta = 804$), BBDC4 ($\Delta = 1006$) and USIM5 ($\Delta = 841$). Results are shown for $R = 10$. The curves suggest that there is no consistent tendency of increasing or decreasing of accuracy when k is increased. This is confirmed by the variation of the sign of the angular coefficients resulting from linear regressions displayed in Table 5.

the limitations of computational speed and communication, and trading fees, such predictability is at least market inefficiency clue.

We applied the model using data on shares of companies in the Brazilian stock exchange BM&FBOVESPA; for each asset, we collected a number of prices sampled every 5 minutes enough to perform 4,500 forecasts. The choice of stocks of an emerging market allowed us to perform this first model test in an instance where predictability is more probable to be observed. For this data, we found through statistical hypothesis tests that the accuracies of the predictions obtained are greater than that of a completely random model in a significant number of cases. The differential equation model produced results statistically different from the

Table 6: Statistics of forecasts in the sets $\bar{V}_4$ for different values of the parameter R.

ITSA4	R	1	2	3	4	5	6	7	8	9	10
$\bar{V}_4$	n	857	844	873	889	878	866	876	879	873	869
	a	0.5309	0.5427	0.5601	0.5703	0.5706	0.5715	0.5719	0.5756	0.5864	0.5903
	p	0.0704	**0.0131**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**
BBDC4	R	1	2	3	4	5	6	7	8	9	10
$\bar{V}_4$	n	1057	1035	1048	1057	1067	1097	1082	1077	1083	1082
	a	0.5260	0.5188	0.5572	0.5506	0.5651	0.5806	0.5878	0.5831	0.5641	0.5730
	p	0.0900	0.2264	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**
USIM5	R	1	2	3	4	5	6	7	8	9	10
$\bar{V}_4$	n	913	882	903	914	925	903	916	913	902	888
	a	0.5334	0.5328	0.4939	0.5164	0.5243	0.5293	0.5436	0.5377	0.5399	0.5360
	p	**0.0435**	0.0513	0.7139	0.3213	0.1393	0.0782	**0.0083**	**0.0227**	**0.0165**	**0.0319**

Note: The symbol n represents the total number of forecasts, a is the accuracy and p is the p-value for hypothesis test assuming as the null hypothesis that the forecast model gives the right answer half of the times in the sets analyzed. Highlighted in bold 22 p-values less than 0.05, corresponding to approximately 73% of the cases. The colored cells correspond to the cases where the sample violates the premise of independence of data (27%).

ones of the fair coin model, i.e., high accuracies or correlated sequences of successes and fails were found, in nearly 59% of the cases in tables (1,2 and 3), which presents results for all subsets $\bar{U}$, $\bar{U}_1$, $\bar{U}_2$, $\bar{U}_3$ and $\bar{U}_4$, and in approximately 83% of the cases in Table 6, where the results refer to $\bar{V}_4$. This indicates the presence of information on future price in the series analyzed.

The accuracies calculated showed a tendency to be higher when the model was adjusted using data covering larger intervals (high values of R). In such situations, the average variability of adjacent matrices decreased, since more data were shared in the calculation of matrices corresponding to adjacent times. This decreased variability promoted the effectiveness of the model. Accuracy was also likely to rise when computed considering only the trends associated with the forecasted returns with high magnitudes (the ones in $\bar{U}_4$ or $\bar{V}_4$). This shows that the absolute values of predicted returns also carry information about future prices. Finally, the presence of linear correlation collaborated to enhance the predictive power of the model. Thus, results statistically different than the ones from a fair coin model was found in all cases regarding the highly correlated stocks ITSA4 and BBDC4 for $R \geq 3$ and forecasts in the sets $\bar{U}_4$ or $\bar{V}_4$ (see tables (1,2, 3 and 6)).

We analyzed the influence of volatility, calculated for each forecast in a period immediately preceding it, on accuracy: no relation was observed. The dependence of accuracy on the number of stocks used in

Table 7: Confusion matrices of forecasts resulting from votings involving the forecasts in the sets $\bar{V}_4$ with R ranging from 6 to 10.

ITSA4		
Actual \ Predicted	High	Low
High	365 (59%)	253 (41%)
Low	279 (45%)	343 (55%)

BBDC4		
Actual \ Predicted	High	Low
High	434 (56%)	335 (44%)
Low	333 (42%)	456 (58%)

USIM5		
Actual \ Predicted	High	Low
High	335 (51%)	321 (49%)
Low	297 (45%)	359 (55%)

the model was also studied. The model was fed with data of 2, 3, 4 and 5 stocks. A clear increase in the predictive power was observed when we change from 2 to 3 assets, but no consistent further gain was found for 4 and 5 stocks. Since the model proposed here is not among the extensively explored in the literature, there are many possibilities for further refinements which may lead to greater accuracies. The introduction of nonlinearities is one of them.

The utilization of the model as part of a trading strategy was investigated taking into account a voting method including forecasts from different versions of it. The accuracies found for this case are close to the ones resulting from the versions that took part in the votings, but a greater number of predictions were provided. According to the evaluation metrics calculated, the performance of the model for *predicted* positive returns is equivalent to its performance for negative ones; the same is valid for *actual* positive and negative returns.

The application of this model to assets of different markets and the use of other sampling intervals, in

Table 8: Evaluation metrics of forecasts resulting from votings involving the forecasts in the sets $\bar{V}_4$ with R ranging from 6 to 10.

Terminology	Formula	ITSA4	BBDC4	USIM5
Accuracy	$\frac{Tp+Tn}{n}$	0.5709	0.5712	0.5289
Error	1-Accuracy	0.4290	0.4288	0.4710
Sensitivity	$\frac{Tp}{Tp+Fn}$	0.5906	0.5643	0.5106
Specificity	$\frac{Tn}{Tn+Fp}$	0.5514	0.5779	0.5472
Positive predictive value	$\frac{Tp}{Tp+Fp}$	0.5667	0.5658	0.5300
Negative predictive value	$\frac{Tn}{Tn+Fn}$	0.5755	0.5764	0.5279
F1 scoreA	$\frac{2Tp}{2Tp+Fp+Fn}$	0.5784	0.5651	0.5201
F1 scoreB	$\frac{2Tn}{2Tn+Fn+Fp}$	0.5632	0.5772	0.5374

Note: The symbol Tp represents the true positive forecasts, Tn the true negative forecasts, Fp the false positive forecasts, Fn the false negative forecasts and n the total number of forecasts.

order to verify if the results found here remain, correspond to future steps of this research. In order to more fully explore collective dynamics, it would be useful to perform forecasts involving larger numbers of shares ($N > 5$). A question to be answered refers to the possibility of improving the model predictive power by means of increasing the intervals where the coefficient matrices are fitted ($R > 10$). A related topic concerns the seeking of values of R beyond which the accuracies decay: for too large intervals, part of the data feeding the model may be non correlated with the forecasts, disturbing the model functioning. New computation tools are necessary for significantly enhance N or R. In view of the study of market efficiency, it is also worthwhile to perform a realistic simulation of a trading strategy based on the model aiming at the effective verification of arbitrage opportunities.

5. Acknowledgements

ARBM acknowledges the support from the Brazilian agencies FAPEMIG (Grant APQ-02027-14), CNPq (Grants 486920/2012-7, 305380/2012-5 and 306871/2012-2) and CEFET/MG (PROPESQ program—Grant 10122_2012). ACMP acknowledges the support of the Brazilian National Institute of Science and Tech-

Table 9: Statistics of forecasts for variable number of stocks in the model.

N		ITUB4	PETR4	VALE5	BBDC4	ABEV3
2	n	1508	1452			
	a	0.5232	0.5179			
3	n	1514	1474	1505		
	a	0.5449	0.5597	0.5395		
4	n	1534	1496	1533	1595	
	a	0.5482	0.5668	0.5407	0.5761	
5	n	1593	1521	1544	1588	1395
	a	0.5417	0.5700	0.5310	0.5761	0.5373

Note: The forecasts whose statistics are shown are the result of votings involving the forecasts in the sets $\bar{V}_4$ with R ranging from 6 to 10. In each line are displayed the results concerning the stocks used to run the model. The symbol N represents the number stocks, n is the number of forecasts and a is the accuracy.

nology for the Web (grant no. 573871/2008-6), MASWeb (grant FAPEMIG/PRONEX APQ-01400-14), CAPES, CNPq (grant 459301/2014-4), Finep, and Fapemig.

References

[1] P. A. Samuelson, Proof that properly anticipated prices fluctuate randomly, 1965.

[2] L. Feng, B. Li, B. Podobnik, T. Preis, H. E. Stanley, Linking agent-based models and stochastic models of financial markets, Proceedings of the National Academy of Sciences 109 (22) (2012) 8388–8393. doi:10.1073/pnas.1205013109.

[3] B. G. Malkiel, E. F. Fama, Efficient capital markets: A review of theory and empirical work, The journal of Finance 25 (2) (1970) 383–417. doi:10.2307/2325486.

[4] R. J. Shiller, Do stock prices move too much to be justified by subsequent changes in dividends? (1980). doi:10.3386/w0456.

[5] L. P. Hansen, K. J. Singleton, Generalized instrumental variables estimation of nonlinear rational expectations models, Econometrica: Journal of the Econometric Society 50 (5) (1982) 1269–1286. doi:10.2307/1911873.

[6] D. Kahneman, A. Tversky, Prospect theory: An analysis of decision under risk, Econometrica: Journal of the econometric society 47 (2) (1979) 263–291. doi:10.2307/1914185.

[7] R. J. Shiller, S. Fischer, B. M. Friedman, Stock prices and social dynamics, Brookings papers on economic activity 1984 (2) (1984) 457–510. doi:10.2307/2534436.

[8] V. L. Smith, An experimental study of competitive market behavior, The Journal of Political Economy 70 (2) (1962) 111–137. doi:http://www.jstor.org/stable/1861810.

[9] G. Caginalp, D. Porter, V. Smith, Financial bubbles: Excess cash, momentum, and incomplete information, The Journal of Psychology and Financial Markets 2 (2) (2001) 80–99. doi:10.1207/S15327760JPFM0202_03.

[10] A. W. Lo, The adaptive markets hypothesis: Market efficiency from an evolutionary perspective, Journal of Portfolio Management, Forthcoming 30 (5) (2004) 15–29. doi:10.3905/jpm.2004.442611.

[11] J. T. Barkoulas, C. F. Baum, Long-term dependence in stock returns, Economics Letters 53 (3) (1996) 253–259. doi:10.1016/S0165-1765(96)00935-4.

[12] D. Grech, Z. Mazur, Can one make any crash prediction in finance using the local hurst exponent idea?, Physica A: Statistical Mechanics and its Applications 336 (1) (2004) 133–145. doi:10.1016/j.physa.2004.01.018.

[13] D. O. Cajueiro, B. M. Tabak, The hurst exponent over time: testing the assertion that emerging markets are becoming more efficient, Physica A: Statistical Mechanics and its Applications 336 (3) (2004) 521–537. doi:10.1016/j.physa.2003.12.031.

[14] L. Zunino, B. M. Tabak, D. G. Pérez, M. Garavaglia, O. A. Rosso, Inefficiency in latin-american market indices, The European Physical Journal B 60 (1) (2007) 111–121. doi:10.1140/epjb/e2007-00316-y.

[15] T. Chordia, R. Roll, A. Subrahmanyam, Liquidity and market efficiency, Journal of Financial Economics 87 (2) (2008) 249–268. doi:10.1016/j.jfineco.2007.03.005.

[16] J. Belaire-Franch, K. K. Opong, Testing for random walk in euro exchange rates using the subsampling approach, Applied Economics Letters 17 (12) (2010) 1145–1151. doi:10.1080/00036840902817581.

[17] G. Caginalp, M. DeSantis, A paradigm for quantitative behavioral finance, American Behavioral Scientist 55 (8) (2011) 1014–1034. doi:10.1177/0002764211412356.

[18] A. Charles, O. Darné, J. H. Kim, Exchange-rate return predictability and the adaptive markets hypothesis: Evidence from major foreign exchange rates, Journal of International Money and Finance 31 (6) (2012) 1607–1626. doi:10.1016/j.jimonfin.2012.03.003.

[19] O. M. Al-Khazali, C. S. Pyun, D. Kim, Are exchange rate movements predictable in asia-pacific markets? evidence of random walk and martingale difference processes, International Review of Economics & Finance 21 (1) (2012) 221–231. doi:10.1016/j.iref.2011.07.002.

[20] D. Lazăr, A. Todea, D. Filip, Martingale difference hypothesis and financial crisis: Empirical evidence from european emerging foreign exchange markets, Economic Systems 36 (3) (2012) 338–350. doi:10.1016/j.ecosys.2012.02.002.

[21] T. Choudhry, R. Jayasekera, Comparison of efficiency characteristics between the banking sectors of us and uk during the global financial crisis of 2007–2011, International Review of Financial Analysis 25 (2012) 106–116. doi:10.1016/j.irfa.2012.09.002.

[22] T. Choudhry, R. Jayasekera, Market efficiency during the global financial crisis: Empirical evidence from european banks, Journal of international Money and Finance 49 (2014) 299–318. doi:10.1016/j.jimonfin.2014.03.008.

[23] L. Katusiime, A. Shamsuddin, F. W. Agbola, Foreign exchange market efficiency and profitability of trading rules: Evidence from a developing country, International Review of Economics & Finance 35 (2015) 315–332. doi:10.1016/j.iref.2014.10.003.

[24] C. G. Ntim, J. English, J. Nwachukwu, Y. Wang, On the efficiency of the global gold markets, International Review of Financial Analysis 41 (2015) 218–236. doi:10.1016/j.irfa.2015.03.013.

[25] C. Curme, H. E. Stanley, I. Vodenska, Coupled network approach to predictability of financial market returns and news sentiments, International Journal of Theoretical and Applied Finance 18 (07) (2015) 1550043. doi:10.1142/S0219024915500430.

[26] B. Podobnik, D. Fu, T. Jagric, I. Grosse, H. E. Stanley, Fractionally integrated process for transition economics, Physica A: Statistical Mechanics and its Applications 362 (2) (2006) 465–470. doi:10.1016/j.physa.2005.09.051.

[27] S. Liu, Transaction costs and market efficiency: Evidence from commission deregulation, The Quarterly Review of Economics and Finance 50 (3) (2010) 352–360. doi:10.1016/j.qref.2010.02.001.

[28] D. Chung, K. Hrazdil, Liquidity and market efficiency: A large sample study, Journal of Banking & Finance 34 (10) (2010) 2346–2357. doi:10.1016/j.jbankfin.2010.02.021.

[29] S. A. Ross, Information and volatility: The no-arbitrage martingale approach to timing and resolution irrelevancy, The Journal of Finance 44 (1) (1989) 1–17. doi:10.2307/2328272.

[30] P. M. Healy, K. G. Palepu, Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature, Journal of accounting and economics 31 (1) (2001) 405–440. doi:10.1016/S0165-4101(01)00018-0.

[31] M. K. Brunnermeier, Information leakage and market efficiency, Review of Financial Studies 18 (2) (2005) 417–457. doi:10.1093/rfs/hhi015.

[32] M. Rafferty, M. Tomljanovich, Central bank transparency and market efficiency: An econometric analysis, Journal of Economics and Finance 26 (2) (2002) 150–161. doi:10.1007/BF02755982.

[33] M. D. Bordo, J. Landon-Lane, Does expansionary monetary policy cause asset price booms; some historical and empirical evidence, Tech. rep., National Bureau of Economic Research (2013).

[34] A. Alonso-Rivera, S. Cruz-Aké, F. Venegas-Martínez, Impact of monetary policy on financial markets efficiency and speculative bubbles: A non-linear entropy-based approach.

[35] P.-w. Chen, H.-c. Huang, Y.-c. Su, The central bank in market efficiency: The case of taiwan, Pacific-Basin Finance Journal 29 (2014) 239–260. doi:10.1016/j.pacfin.2014.04.002.

[36] P. Pasquariello, C. Zafeiridou, Political uncertainty and financial market quality, Ross School of Business Paper (1232). doi:10.2139/ssrn.2423576.

[37] K.-P. Lim, R. D. Brooks, J. H. Kim, Financial crisis and stock market efficiency: Empirical evidence from asian countries, International Review of Financial Analysis 17 (3) (2008) 571–591. doi:10.1016/j.irfa.2007.03.001.

[38] T. Preis, D. Y. Kenett, H. E. Stanley, D. Helbing, E. Ben-Jacob, Quantifying the behavior of stock correlations under market stress, Scientific reports 2 (752). doi:10.1038/srep00752.

[39] M. Karanasos, S. Yfanti, M. Karoglou, Multivariate fiaparch modelling of financial markets with dynamic correlations in times of crisis, International Review of Financial Analysis 45 (2014) 332â349. doi:10.1016/j.irfa.2014.09.002.

[40] X. Jin, X. An, Global financial crisis and emerging stock market contagion: A volatility impulse response function approach, Research in International Business and Finance 36 (2016) 179–195. doi:10.1016/j.ribaf.2015.09.019.

[41] L. Yarovaya, M. C. K. Lau, Stock market comovements around the global financial crisis: Evidence from the uk, brics and mist markets, Research in International Business and Finance 37 (2016) 605–619. doi:10.1016/j.ribaf.2016.01.023.

[42] D. Y. Kenett, M. Raddant, T. Lux, E. Ben-Jacob, Evolvement of uniformity and volatility in the stressed global financial village, PloS one 7 (2) (2012) e31144. doi:10.1371/journal.pone.0031144.

[43] D. Harmon, M. Lagi, M. A. de Aguiar, D. D. Chinellato, D. Braha, I. R. Epstein, Y. Bar-Yam, Anticipating economic market crises using measures of collective panic, PloS one 10 (7) (2015) e0131871. doi:10.1371/journal.pone.0131871.

[44] M. Rechenthin, W. N. Street, Using conditional probability to identify trends in intra-day high-frequency equity pricing, Physica A: Statistical Mechanics and its Applications 392 (24) (2013) 6169–6188. doi:10.1016/j.physa.2013.08.003.

[45] T. Preis, H. S. Moat, H. E. Stanley, Quantifying trading behavior in financial markets using google trends, Scientific reports 3. doi:10.1038/srep01684.

[46] H. S. Moat, C. Curme, A. Avakian, D. Y. Kenett, H. E. Stanley, T. Preis, Quantifying wikipedia usage patterns before stock market moves, Scientific reports 3. doi:10.1038/srep01801.

[47] G. S. Atsalakis, K. P. Valavanis, Surveying stock market forecasting techniques–part ii: Soft computing methods, Expert Systems with Applications 36 (3) (2009) 5932–5941.

[48] G. S. Atsalakis, K. P. Valavanis, Surveying stock market forecasting techniques–part ii: Soft computing methods, Expert Systems with Applications 36 (3) (2009) 5932–5941. doi:10.1016/j.eswa.2008.07.006.

[49] T. Bury, Predicting trend reversals using market instantaneous state, Physica A: Statistical Mechanics and its Applications 404 (2014) 79–91. doi:10.1016/j.physa.2014.02.044.

[50] T. W. Epps, Comovements in stock prices in the very short run, Journal of the American Statistical Association 74 (366a) (1979) 291–298. doi:10.2307/2286325.

[51] E. C. Kaplanis, Stability and forecasting of the comovement measures of international stock market returns, Journal of International Money and Finance 7 (1) (1988) 63–75. doi:10.1016/0261-5606(88)90006-X.

[52] S. Lee, The stability of the co-movements between real estate returns in the uk, Journal of Property Investment & Finance 24 (5) (2006) 434–442. doi:10.1108/14635780610691913.

[53] T. D. Peron, F. A. Rodrigues, Collective behavior in financial markets, EPL (Europhysics Letters) 96 (4) (2011) 48004. doi:10.1209/0295-5075/96/48004.

[54] S.-H. Poon, C. W. Granger, Forecasting volatility in financial markets: A review, Journal of economic literature 41 (2) (2003) 478–539. doi:10.1257/002205103765762743.

[55] E. Hajizadeh, A. Seifi, M. F. Zarandi, I. Turksen, A hybrid modeling approach for forecasting the volatility of s&p 500 index return, Expert Systems with Applications 39 (1) (2012) 431–436. doi:10.1016/j.eswa.2011.07.033.

[56] W. Kristjanpoller, A. Fadic, M. C. Minutolo, Volatility forecast using hybrid neural network models, Expert Systems with Applications 41 (5) (2014) 2437–2442. doi:10.1016/j.eswa.2013.09.043.

[57] R. Dash, P. Dash, An evolutionary hybrid fuzzy computationally efficient egarch model for volatility prediction, Applied Soft Computing 45 (2016) 40–60. doi:10.1016/j.asoc.2016.04.014.

[58] H. Nyberg, Forecasting the direction of the us stock market with dynamic binary probit models, International Journal of Forecasting 27 (2) (2011) 561–578. doi:10.1016/j.ijforecast.2010.02.008.

[59] S. Anatolyev, N. Gospodinov, Modeling financial return dynamics via decomposition., Journal of Business & Economic Statisticsdoi:10.1198/jbes.2010.07017.

[60] T. Chevapatrakul, Return sign forecasts based on conditional risk: Evidence from the uk stock market index., Journal of Banking & Finance 37 (7) (2013) 2342–2353. doi:10.1016/j.jbankfin.2013.01.033.

[61] H. Pönkä, Predicting the direction of us stock markets using industry returns, Empirical Economics (2014) 1–30doi:10.1007/s00181-016-1098-0.

[62] P. F. Christoffersen, F. X. Diebold, Financial asset returns, direction-of-change forecasting, and volatility dynamics, Management Science 52 (8) (2006) 1273–1287. doi:10.1287/mnsc.1060.0520.

[63] T. Preis, D. Reith, H. E. Stanley, Complex dynamics of our economic life on different scales: insights from search engine query data, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 368 (1933) (2010) 5707–5719. doi:10.1098/rsta.2010.0284.

[64] E. J. Ruiz, V. Hristidis, C. Castillo, A. Gionis, A. Jaimes, Correlating financial time series with micro-blogging activity, in: Proceedings of the fifth ACM international conference on Web search and data mining, ACM, 2012, pp. 513–522. doi:10.1145/2124295.2124358.

[65] A.-S. Chen, M. T. Leung, H. Daouk, Application of neural networks to an emerging financial market: forecasting and trading the taiwan stock index, Computers & Operations Research 30 (6) (2003) 901–923. doi:10.1016/S0305-0548(02)00037-0.

[66] M. Khashei, S. R. Hejazi, M. Bijari, A new hybrid artificial neural networks and fuzzy regression model for time series forecasting, Fuzzy Sets and Systems 159 (7) (2008) 769–786. doi:10.1016/j.fss.2007.10.011.

[67] L. Yu, H. Chen, S. Wang, K. K. Lai, Evolving least squares support vector machines for stock market trend mining, IEEE Transactions on evolutionary computation 13 (1) (2009) 87–102. doi:10.1109/TEVC.2008.928176.

[68] G. S. Atsalakis, K. P. Valavanis, Forecasting stock market short-term trends using a neuro-fuzzy based methodology, Expert Systems with Applications 36 (7) (2009) 10696–10707. doi:10.1016/j.eswa.2009.02.043.

[69] Y. Kara, M. A. Boyacioglu, Ö. K. Baykan, Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the istanbul stock exchange, Expert systems with Applications 38 (5) (2011) 5311–5319. doi:10.1016/j.eswa.2010.10.027.

[70] R. Allez, J.-P. Bouchaud, Individual and collective stock dynamics: intra-day seasonalities, New Journal of Physics 13 (2) (2011) 025010. doi:10.1088/1367-2630/13/2/025010.

[71] S. Ryu, W. Cai, A. Caro, Quantum entanglement of formation between qudits, Physical Review A 77 (5) (2008) 052312. doi:10.1103/PhysRevA.77.052312.

[72] D. C. Montgomery, Design and analysis of experiments, John Wiley & Sons, 2008.