Efficient Market Hypothesis in Brazilian and American Markets: A Study Using Time Windows

Rodrigo Coimbra P. A.[1], Arthur R. B. de Magalhães[1], Charlene C. de Resende[2]

[1]*Postgraduate Program in Mathematical and Computational Modeling, Federal Center of Technological Education of Minas Gerais*
Av. Amazonas 7675 - Nova Gameleira - Belo Horizonte, 30510-000, Minas Gerais, Brazil
rodrigocpalves@hotmail.com
[2]*Department of Mathematics, Federal Center of Technological Education of Minas Gerais*
Av. Amazonas 7675 - Nova Gameleira - Belo Horizonte, 30510-000, Minas Gerais, Brazil
magalhaes@cefetmg.br
[3]*Una University Center*
Av. Maria da Glória Rocha 175 - Bitacula, Contagem, 32010-375, Minas Gerais, Brazil
charlene.resende@animaeducacao.com.br

Abstract. Financial markets play a fundamental role in contemporary societies, directly influencing social well-being. Traditionally, economists have sought to understand the underlying processes driving the dynamics of these markets. However, in recent decades, there has been a significant increase in the contribution of researchers from various disciplines, including physics, mathematics, and computer science. In this context, econophysics emerges, employing typical physics methods to study systems commonly investigated in economics and finance. Financial markets are recognized as complex systems, whose behavior can be observed through indicators such as prices and trading volumes. However, the statistical characterization of these time series remains an ongoing challenge, as does the understanding of their emergence from microeconomic relationships. The Efficient Market Hypothesis (EMH) suggests that time series of future prices in financial markets contain little useful information for prediction, making it extremely difficult. In this study, we conducted analyses of financial series to investigate this expectation. We used a model for forecasting return trends based on differential equations with a focus on segmentation by time windows. The accuracy of this model can provide insights into the presence of information suggested by the EMH in financial series.

Keywords: Price Trends, Efficient Market Hypothesis, Financial Series.

1 Introduction

Humanity has sought to understand market dynamics for millennia. The understanding of the market concept, along with the concept of the free market, has gone through various moments and contexts. Its comprehension has been shaped over time by multiple ideas and viewpoints. Thus, current interpretations of its behaviors and nature reflect a complex historical evolution, a construct that manifests itself in diverse ways in different periods. Understanding the historical category of the market remains a challenging task to this day and will certainly continue to be for an indefinite period.

In the study of financial series, we usually engage in the task of trying to predict future behaviors based on available historical data. Armed with statistical tools and mathematical models, we seek, through econometric analyses, to find trends that assist us in better understanding economic phenomena and behaviors capable of inspiring investment strategies, altering our perception of the very nature of markets, or potentially influencing policies of varying significance.

Aware of these possibilities, in this work, we use a model for predicting price trends based on coupled differential equations capable of analyzing financial series and providing information on price trends of multiple assets at each new instant of time within the same market. In this study, we focus on applying the model to segment the assets studied by time windows. In this way, the aim is to identify the best times of the day to trade the assets that make up the indices covered in the study in order to make profits while minimizing risks. The very idea of achieving this feat, in turn, contradicts the Efficient Market Hypothesis, so our objective becomes to find potential deviations from this hypothesis.

In section 2, we describe the mathematical model used for analyzing the historical series data, in section 3, the findings are presented, and discussions of the results are provided. Finally, in section 4, we present the final discussions and conclusions of the study.

2 Forecast Model

A model for predicting price trends based on systems of differential equations is presented in Fonseca [1]. It allows us to investigate a set of assets simultaneously.

An exponential moving average is used to obtain reference prices for each time instant of each of the evaluated assets. The reference price functions here as a substitute for the fair price of the asset. The deviation from the reference price is given by

$$X_s(t) = \ln x_s(t) - \ln \overline{x}_s(t), \qquad s = 1,2,\cdots,N, \tag{1}$$

where $x_s(t)$ is the closing price of candles with interval Δt and $\overline{x}_s(t)$ is its respective mean calculated over a preceding time window of size 1000 and the first preceding time window has size 12 and is obtained through simple arithmetic mean. The choice of logarithmic scale helps reduce scale problems. It reveals information about the price variation relative to the reference price, indicating whether a stock is high or low compared to the considered fair price.

The model used originates from the differential equation

$$\frac{d}{dt}\mathsf{X}(t) = \mathsf{A}\mathsf{X}(t), \tag{2}$$

where A is a generic coefficients matrix and

$$\mathsf{X}(t) = \begin{pmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ X_N(t) \end{pmatrix}. \tag{3}$$

Equation (2) can be discretized as follows:

$$\frac{\mathsf{X}(t) - \mathsf{X}(t-\Delta t)}{\Delta t} = \mathsf{A}\mathsf{X}(t), \tag{4}$$

which is equivalent to

$$\mathsf{X}(t) = \mathsf{C}\mathsf{X}(t-\Delta t) \tag{5}$$

and

$$\mathsf{C} = (\mathsf{I} - \Delta t\mathsf{A})^{-1}, \tag{6}$$

where I corresponds to the $N \times N$ identity matrix.

We will assume that $\Delta t = 1$, which corresponds to the choice of the time scale. Assuming that the coefficients matrix C exists, we want to determine it from the relationship expressed in eq. (2). Given that C has N^2 elements and eq. (2) presents only N equations, a set of N equations similar to eq. (2) is necessary to balance the number of equations with the number of parameters to be determined. It follows, therefore, that

$$\mathsf{X}_N(t) = \mathsf{C}\mathsf{X}_N(t-1), \tag{7}$$

where

$$\mathsf{X}_N(t) = \begin{pmatrix} X_1(t-N+1) & \cdots & X_1(t-1) & X_1(t) \\ X_2(t-N+1) & \cdots & X_2(t-1) & X_2(t) \\ \vdots & \ddots & \vdots & \vdots \\ X_N(t-N+1) & \cdots & X_N(t-1) & X_N(t) \end{pmatrix}. \tag{8}$$

If $\mathrm{X}_N(t-1)$ is invertible, we will have, for each time t_0

$$\mathsf{C}(t_0) = \mathsf{X}_N(t_0)\mathsf{X}_N^{-1}(t_0-1). \tag{9}$$

In order to find more stable values for $\mathrm{C}(t_0)$, we apply eq. (2) for $M > N$ adjacent time steps, for t from t_0 to $t_0 - M + 1$. Using the least squares method, we arrive at the estimator

$$\hat{\mathsf{C}}(t_0) = \left(\mathsf{X}_M(t_0)\mathsf{X}_M^T(t_0-1)\right)\left(\mathsf{X}_M(t_0-1)\mathsf{X}_M^T(t_0-1)\right)^{-1}, \tag{10}$$

where

$$\mathsf{X}_M(\tau) = \begin{pmatrix} X_1(\tau-M+1) & \cdots & X_1(\tau-1) & X_1(\tau) \\ X_2(\tau-M+1) & \cdots & X_2(\tau-1) & X_2(\tau) \\ \vdots & \ddots & \vdots & \vdots \\ X_N(\tau-M+1) & \cdots & X_N(\tau-1) & X_N(\tau) \end{pmatrix} \tag{11}$$

is a matrix of size $N \times M$. We will use $\widehat{\mathrm{C}}(t_0)$ to estimate the deviations $\hat{X}_s(t_0+1)$ according to

$$\begin{pmatrix} \hat{X}_1(t_0+1) \\ \hat{X}_2(t_0+1) \\ \vdots \\ \hat{X}_N(t_0+1) \end{pmatrix} = \hat{\mathsf{X}}(t_0+1) = \hat{\mathsf{C}}(t_0)\mathsf{X}(t_0). \tag{12}$$

These deviations will be converted into predicted prices and compared with the actual prices, providing information about their trends. Thus, from the estimated reference price, we obtain the asset price at $t_0 + 1$ as follows:

$$\hat{X}_s(t+1) = \ln\hat{x}_s(t+1) - \ln\overline{x}_s(t), \tag{13}$$

which leads to

$$\hat{x}_s(t+1) = \overline{x}_s(t)e^{\hat{X}_s(t+1)}, \tag{14}$$

where $\hat{x}_s(t+1)$ is the predicted price at time $t_0 + 1$ and $\hat{X}_s(t+1)$ is the predicted deviation for the respective time step. The accuracy of the model is determined by the proportion of correct trend predictions n_c (when the predicted and actual variations have the same direction) to the total predictions n made, expressed as

$$a = 100\frac{n_c}{n}, \tag{15}$$

presented here as a percentage.

When the absolute value of the actual variation is zero, the prediction associated with that time step is not considered in the accuracy computation. Additionally, a prediction is not considered valid if its absolute value is not greater than the upper quartile of the predicted values within a window of immediately preceding predictions (the size of this window is one of the model parameters). In other words, for a prediction to be considered valid, the absolute value of its return must be greater than 75% of the predicted returns from the preceding window. We believe that the size of the predicted return enhances the accuracy of the model and allows for more robust results, as noted in Resende [2].

3 Results

In this work, we are using 58 assets that compose the Ibovespa index and 30 assets that compose the Dow Jones index. The analysis of these assets is based on 5-minute candles, which are further grouped by time slots ranging from 10 to 17 hours (starting time in BRT time zone) and each time slot has a duration of 1 hour. To avoid anomalies, we chose to start the time slot from 10 AM to 10:30 AM. We believe that eliminating the initial candles from the price series reduces the effects of the "candle widening" resulting from the interval between the closing of the stock exchange on the previous day and its opening on the following day.

Accuracies were obtained per hour and per asset for the Ibovespa and Dow Jones indices. Figure 1 shows the results for the Ibovespa index assets and Fig. 2 shows the results for the Dow Jones index assets.

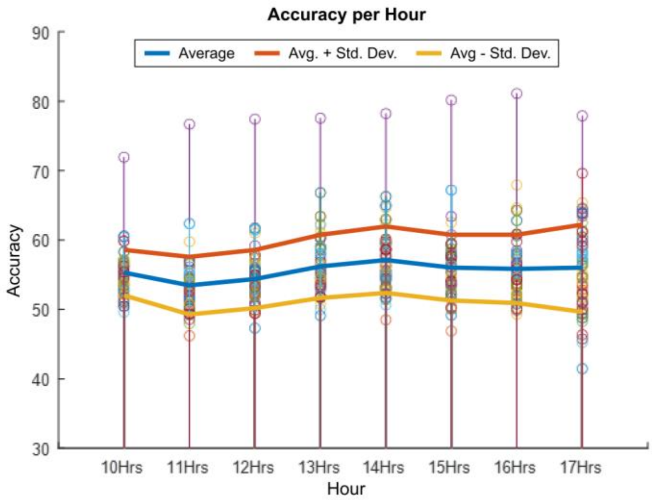

Figure 1. Accuracy per hour for the assets of the Ibovespa index. Each differently colored point corresponds to an action. The lines connecting the hours represent the averages of the accuracies at each hour and their respective standard deviations

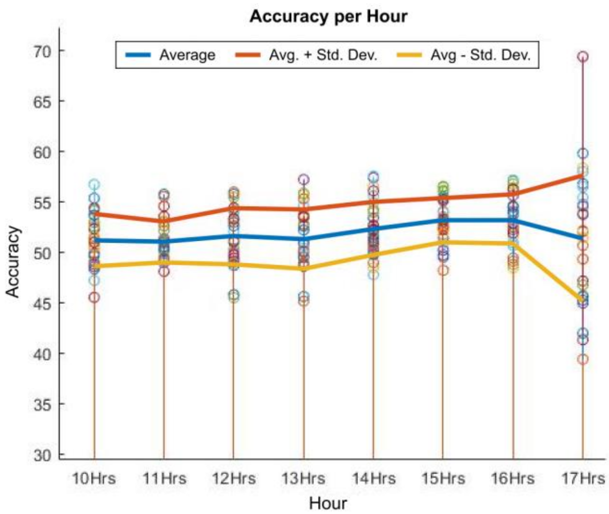

Figure 2. Accuracy per hour for the assets of the Dow Jones index. Each differently colored point corresponds to an action. The lines connecting the hours represent the averages of the accuracies at each hour and their respective standard deviations

On average, the highest accuracies were recorded for the actions that compose both indices in the following time slots: 14, 17, 13, 16, and 15 hours, in descending order. Analyzing the results for the gross average return charts, we present in Fig. 3 the results for the Ibovespa index assets.

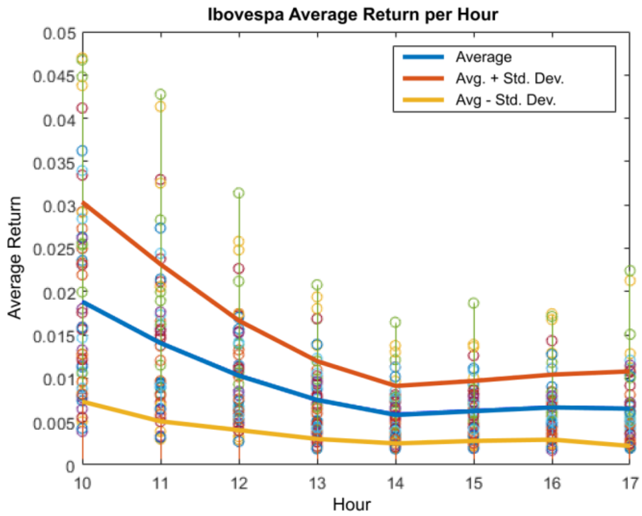

Figure 3. Gross average return per hour for the assets of the Ibovespa index. Each differently colored point corresponds to an action. The lines connecting the hours represent the averages of the returns at each hour and their respective standard deviations

The highest averages for the gross average returns, for both actions, are found in the time windows of 10, 11, and 12 hours, which contrasts with the results obtained for accuracy, as expected by the EMH, where the gross average return would decrease with increasing accuracy and vice versa, indicating the impossibility of arbitrage.

Figure 4 show the results for the Dow Jones index assets.

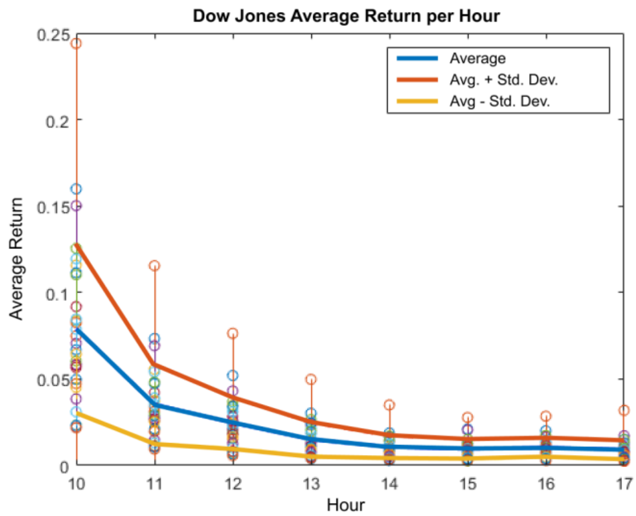

Figure 4. Gross average return per hour for the assets of the Dow Jones index. Each differently colored point corresponds to an action. The lines connecting the hours represent the averages of the returns at each hour and their respective standard deviations

With the results obtained for the analysis of accuracies filtered by time slots and asset, we seek to identify correlations between the gross average returns and the accuracies found in this way. Table (1) displays the Pearson correlation coefficients4, as well as the corresponding p-values.

Table 1. Pearson Coefficients Accuracy versus Average Return for the Ibovespa (B3) and Dow Jones (DJ) indices from 10 to 17 hours

-	10h	11h	12h	13h	14h	15h	16h	17h
Pearson (B3)	-0.26	-0.35	-0.31	-0.44	-0.37	-0.26	-0.38	-0.13
P-value (B3)	0.0458	0.0064	0.0191	0.0005	0.0045	0.0451	0.0032	0.3198
Pearson (DJ)	-0.29	0.11	-0.35	-0.34	-0.09	-0.04	0.06	-0.11
P-vlaue (DJ)	0.1245	0.5751	0.0607	0.0675	0.6300	0.8352	0.7594	0.5682

For the assets of the Ibovespa index, these results indicate a moderate correlation for the indices at all hours, except for 17 hours, where the results are statistically inconclusive, as evidenced by the corresponding p-value. This lack of conclusion can be attributed to the fact that it is an auction time, in which buying and selling dynamics do not follow the conventional pattern. Additionally, we observe negative correlations, as expected, since they point to the inverse relationship between the accuracy of the prediction model and its returns. This reinforces the Efficient Market Hypothesis, as it states that it is not possible to consistently invest to obtain large profits without incurring proportional risks. Lower correlations, in absolute terms, suggest arbitrage opportunities, indicating that an increase in accuracy is less compensated by a decrease in return, which may indicate deviations from market efficiency in associated hours. For the assets of the Dow Jones index, we did not find statistically significant p-values (at a significance level of 5%), which prevents us from conducting a more detailed analysis. However, the lack of statistical significance may reflect the nature of the US market, which, unlike the Brazilian market, is more developed. According to Cajueiro and Tabak [3], in more developed markets, arbitrage opportunities tend to be smaller.

4 Conclusions

The aim of this study is to investigate the efficiency of the Brazilian and US markets through a model based on systems of differential equations for price trend prediction. Thus, we seek to deepen our understanding of price dynamics in the financial market, potentially contributing to a broader understanding of the evolution of the Brazilian and US markets towards efficiency, as expected by the Efficient Market Hypothesis (EMH).

Several authors argue that price series contain information (correlations) over time, enabling the identification of trends, which may indicate market inefficiencies. With this in mind, we focus our investigations on the Brazilian and US markets, aiming to compare the characteristics of an emerging market (less efficient) with those of a developed market (more efficient).

Using hourly intervals starting at 10, 11, 12, 13, 14, 15, 16, and 17 hours, we calculated the accuracies and average returns of each asset from both indices. For the Ibovespa, it was observed that, in general, stocks with higher accuracies at a certain time tend to have lower returns at the same time. This result is aligned with the EMH, as it suggests a trade-off between accuracy and average return size regarding arbitrage opportunities. Additionally, we found that the correlations between accuracies and average returns per hour in the Dow Jones index are not statistically significant. In contrast, for the B3, these correlations are significant and may indicate hours of greater or lesser deviation from market efficiency. We noticed that the 13 o'clock hour is the most efficient, suggesting lower arbitrage opportunities, while the 17 o'clock hour is the least efficient, indicating higher arbitrage opportunities.

The formal identification of arbitrage opportunities is a challenging task due to the complexity and variability of the involved agents, whose behaviors are influenced by a combination of personal, political, and social factors. The study of financial markets, particularly concerning the EMH, is therefore a challenging and multifaceted area. Expanding our understanding of the dynamics governing the financial market is a fascinating endeavor that offers numerous possibilities for improving both analytical methods and interpretations of collected data.

Acknowledgements. We thank the Coordination of Superior Level Staff Improvement (CAPES) and the Research Support Foundation of the State of Minas Gerais (FAPEMIG) for their financial support.

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] Fonseca, C. L. G. et al. Is the choice of the candlestick dimension relevant in econophysics? Physica A: Statistical Mechanics and its Applications, Elsevier, v. 582, p. 126233, 2021.
[2] Resende et al., C. C. Investigating market efficiency through a forecasting model based on differential equations. Physica A: Statistical Mechanics and its Applications, Elsevier, v. 474, n. C, p. 199–212, 2017.
[3] Cajueiro, D. O.; Tabak, B. M. The hurst exponent over time: testing the assertion that emerging markets are becoming more efficient. Physica A: Statistical Mechanics and its Applications, Elsevier, v. 336, n. 3, p. 521–537, 2004.