Combining an LSTM neural network with the Variance Ratio Test for time series prediction and operation on the Brazilian stock market

1st Caio Mário Mesquita
Department of Computer Science
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil
caioboninho@gmail.com

2nd Renato Arantes de Oliveira
Department of Computer Science
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil
ra.oliveira@dcc.ufmg.br

3rd Adriano César Machado Pereira
Department of Computer Science
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil
adrianoc@dcc.ufmg.br

***Abstract*—Forecasting financial time series is a problem studied by researchers from different fields, who are looking for effective ways to achieve financial gains. Over time, many authors conducted studies on the possible predictability of the series through different statistical tests, and recently several papers explore the application of machine learning algorithms to have better predictions. In this paper we analyzed real data of 11 time series related to Brazilian stocks, focusing on the statistical characteristics of the series and the use of an LSTM neural network to classify future values. We analyzed the results of 5 different variance ratio tests and their relationship with the neural network classification performance. This paper proposes the application of statistical tests in the LSTM training set to highlight previously those series that have more temporal dependence and, therefore, possibly better forecast results. The results showed that 5 out of 11 stocks rejected the random walk hypothesis through the variance ratio tests and that these same stocks obtained the best performances in terms of classification and financial return.**

***Index Terms*—Financial Market, Machine Learning, LSTM Neural Network, Variance Ratio Test, Algotrading**

I. Introduction

The financial market is intrinsically related to a country's economy. It is responsible for moving a huge amount of money daily. Many economists, investors, researchers, and academics study the market to understand its behavior in an attempt to get a reasonable forecast of financial stocks. This problem has been studied for over a century [1] and understanding how financial market stocks behave is a challenge.

Over the years different statistical tests were developed to test the hypothesis that the financial series followed a random walk. Showing that the series does not follow a random walk is evidence that there may be a temporal dependence on them and justifies the use of some prediction strategy. One important test is the variance ratio, originally proposed in [2]. Then, other tests based on the same concept were developed [3], [4], [5]. Several studies have been done seeking to confront the Random Walk Hypothesis (RWH) in the financial series through these tests, for example, the papers of [6] and [7] in the Asian market.

With the advancement of computational power, many machine learning models have recently been used to capture the behavior of market stocks. A considerable number of papers using the Long Short-Term Memory (LSTM) network can be cited to predict trends in the prices of financial series. The authors of [8] used a combination of LSTM network together with technical market indicators to make forecasts on the Brazilian stock exchange. The results found were promising, with Accuracy values of up to 55.9%. [9] implemented a Bayesian LSTM network model using six indicators from the Chinese market. The results demonstrated that the proposed model increased the results of the original neural network by about 25%. [10] used a genetic algorithm to optimize the internal weights of the network LSTM and demonstrated to reduce the network training time with the technique. [11] compared the performance of an LSTM network with several algorithms (Support Vector Machine, Random Forest, regression) and highlighted the better performance of the LSTM network compared to other techniques using financial series. The comparison of the LSTM architecture with other neural networks also is carried out in [12] on the Sri Lanka market, and the authors highlight the best results in Accuracy terms for the LSTM neural network. We also cite other papers that used LSTM neural networks combined with different data mining and optimization techniques to improve the performance of the forecast ([13], [14], [15], [16]).

All previous cited papers focus either on the series analysis or the direct application of prediction algorithms. In our literary survey, we did not find any work that related the two areas seeking to find direct relationships between the statistical tests in the series and the results of the forecasts with machine learning. The objective of this paper is to apply an LSTM neural network for forecasting Brazilian financial time series and to analyze the relationship between forecasting performance and the results of different variance ratio tests in the series. The hypothesis raised in this paper suggests that it

This work was partially funded by the Brazilian National Institute of Science and Technology for the Web (grant no. 573871/2008-6), MASWeb (grant FAPEMIG/PRONEX APQ-01400-14), CAPES, CNPq (grant 459301/2014-4), Finep, and Fapemig.

is possible to select some series using these statistical tests to apply the machine learning algorithm and obtain better results than in the other series that do not reject the RWH. So, if the training set series of the machine learning algorithm rejects the RWH, then the chance of getting better results is higher.

The remainder of this paper is organized as follows: Section 2 presents the theoretical background addressing the variance ratio test and the LSTM neural network. The following is the methodology adopted in Section 3. Section 4 presents the experiments performed and the results found. Finally, section 5 describes the conclusions and future work.

II. Theoretical foundation

This section presents the main techniques used to develop the work. It is divided into two parts: analysis of significant correlations on the series using variance ratio tests, and the use of an LSTM neural network to classify future values.

A. Variance Ratio Test

Comparisons between the variance of returns of a period with the sum of multiple periods are used to test the RWH. Several tests attempt to exploit any deviations from this prediction, one of the most famous being the variance ratio test proposed by [2]. The test is based on the property that if the series of financial returns follows a random walk, the variance of the sum of consecutive returns must be equal the sum of the individual variances. Following the same development as [17], the intuition behind the test is described below.

Assuming that the return-generating process is stationary with a period variance $V(1) = var(r_t)$. The return of two periods is the sum of consecutive periods and their variance is equal to:

$$\begin{aligned} V(2) = var(r_t + r_{t+1}) = var(r_t) + var(r_{t+1}) \\ +2cov(r_t, r_{t+1}) = (2 + 2\rho_1)V(1), \end{aligned} \tag{1}$$

with ρ_1 being the value of the first auto-correlation lag of the serie. The ratio of the two periods is then defined as:

$$VR(2) = \frac{V(2)}{2V(1)} = 1 + \rho_1. \tag{2}$$

The auto-correlation term is zero when RWH applies and then the variance ratio is 1. Otherwise, the hypothesis of RWH is false and the ratio may be either greater or less than 1.

Considering a period of N returns, where N is an integer greater than or equal to 2. When the RWH hypothesis is true,

$$\begin{aligned} V(N) = var(r_t + r_{t+1} + ... + r_{t+N-1}) \\ = var(r_t) + var(r_{t1}) + ... + var(r_{t+N-1}) = NV(1) \end{aligned} \tag{3}$$

and so the variance is 1 for every N:

$$VR(N) = \frac{V(N)}{NV(1)} = 1. \tag{4}$$

When the hypothesis of RWH is false, V (N) equals $NV(1)$ plus the covariance terms between all distinct return pairs, thus:

$$V(N) = NV(1) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} cov(r_{t+i-1}, r_{t+j1}) \tag{5}$$

$$V(N) = V(1)[N + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \rho_{j-i}] \tag{6}$$

$$VR(N) = 1 + \frac{2}{N}\sum_{\tau=1}^{N-1}(N-\tau)\rho_\tau. \tag{7}$$

The empirical test uses the observed returns to decide if the estimated sample variance ratio is compatible with the theoretical prediction of 1. The test rejects the RWH hypothesis when the ratio value is different from 1. This happens when a function $(N-1)$ first auto-correlations

$$(N-1)\rho_1 + (N-2)\rho_2 + (N-3)\rho_3 + ... + 2\rho_{N-2} + \rho_{N-1}, \tag{8}$$

distance from zero.

This property was explored in two tests M1 and M2 developed by [2], to verify the RWH under the assumption of independently and identically distributed (iid) and conditional heteroscedasticity of the time series, respectively. These tests follow a normal distribution asymptotically. In [4] the author develops 4 alternative non-parametric tests of variance ratio using the rankings (R1, R2) and signals (S1, S2) of the series. Some of the advantages of these tests over the previous are that they may be more powerful than other tests if the data are highly non-normal, and there is no need to resort to asymptotic approximation of the sampling distribution. The tests based on the rankings of the series are accurate under the assumption of i.i.d., whereas the tests based on the signals are accurate under the assumption of conditional heteroscedasticity.

In this paper, we used the statistics M1, M2, R1, R2, and S1. The critical values of the R1, R2 and S1 statistics can be obtained by simulating their exact sample distribution.

B. LSTM neural network

Long-Short Term Memory (LSTM) neural networks were proposed by [18] and are a type of recurrent neural network of deep learning. Recurrent networks have cross-processing feedback connections between past network processing and present-time inputs, thus obtaining the characteristic of memory. This memory property is used to find correlations between separate events that occurred in different temporal moments. For this reason, these networks are ideal for tasks of classification, processing, and prediction in time series of data.

The challenge of using usual recurrent networks to process long data sequences is that they suffer from a problem known in the literature as gradient vanishing. When using backpropagation-based training methods, the error gradient update of farther layer weights may not happen, in which case the network may interrupt its learning process. LSTM networks have been used because it offers better performances in that case by using ports (*input gate*, *forget gate*, *output gate*) that are capable of discarding, maintaining, add or update information on time. These ports – also called gates– help in the keeping the gradient of the error which can be propagated through time and deeper layers of the network. By keeping the gradient of error of deeper layers, the network is able to keep on learning to associate previous states to current inputs in order to make the best prediction and thus reducing the error.

The *forget gate* is responsible for deciding how much information should be discarded when comparing the inputs of the current moment x_t with the values of the past state h_{t-1}. A sigmoid function is used in the output which corresponds to real numbers in the range $[0, 1]$ indicating how much information should flow. The *input gate* is responsible for choosing the values to be updated, and a *tanh* layer creates a vector of new candidates that can be added to the current state. The *output gate* is responsible for providing outputs based on the current state. A combination of a sigmoid function and a *tanh* layer is used to filter the output. Figure 1 illustrates the components of an LSTM unit.

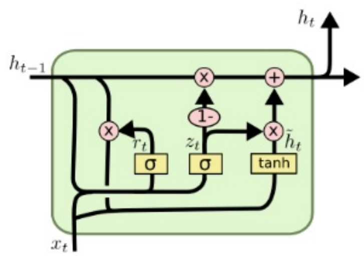

Fig. 1. Representation of an LSTM unit with its respective ports, inputs and outputs. Figure of the work by [19]

III. Methodology

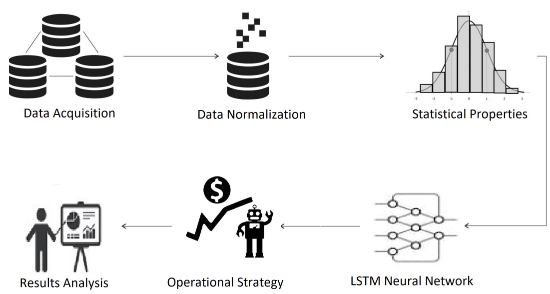

Fig. 2. Work methodology steps

This section describes the methodology developed for applying the strategy in the financial market. The six steps consist of data acquisition, data normalization, statistical properties of the series, LSTM network, operation strategy, and results analysis. Figure 2 illustrates the steps followed in the work methodology.

A. Data Acquisition

It was used the historical data of daily closing prices related to the Brazilian stock series of 2016[1], with a total of 250 days. The data used were the closing price series of 11 stocks: AmBev (ABEV3), Banco do Brasil SA (BBAS3), Braskem (BRKM5), CPFL Energia (CPFE3), Engie Brasil (EGIE3), Eletrobras (ELET3), JBS (JBSS3), Lojas Renner (LREN3), Multiplan (MULT3), Telefônica Brasil (VIVT4) and WEG S.A. (WEGE3). These stocks were chosen because they represent different sectors of the economy (non-cyclical consumption, cyclical consumption, basic materials, public utility, energy, financial, telecomunication and industrial goods) and have a large associated financial volume. It is worth mentioning that the majority of the stocks were part of the main Brazilian index (BOVA11) in 2016.

Data were separated into a 166-day training set and an 84-day test set in the classification step through the LSTM neural network.

B. Data normalization

The data collected has been normalized to be able to work with financial return values rather than daily stock prices. According to [17], the statistical analysis directly applied to the original prices is more difficult as consecutive prices are highly correlated and the price variance increases over time. Prices are not stationary and consequently, it becomes more convenient to analyze the price variation. Results obtained for the variation can easily be used to provide results at the original price.

Thus, the price variation (return) used is given by:

$$r_t = log(p_t) - log(p_{t-1}) \tag{9}$$

where p_t is the price on day t.

C. Statistical properties of series

The training set was used to perform the analysis of the statistical properties of the series. The first 4 statistical moments of the series (mean, standard deviation, skewness, and kurtosis) and the Jarque-Bera test were analyzed to verify the possible normal distribution of the series.

Then 5 different variance ratio tests were applied: M1, M2, R1, R2, and S1. The idea is to verify if there is a relationship between the performance of the machine learning algorithm and the stocks that reject the RWH..

In this step the software *R* was used along with the *moments* and *vrtest* packages.

D. LSTM neural network

This step is responsible for using an LSTM neural network to predict an upward or downward trend in time series observations. We can model a classification problem with two classes: class High, corresponds to the upward movement in the series of returns and class Low corresponds to the downward trend. Class High is assigned to all values of returns greater than or equal to zero, and class Low is assigned to all values less than zero.

The year of 2016 was chosen to apply the prediction model of the neural network. The training set corresponds to two thirds (approximately the months of January to August) with 166 days. The data set corresponds to the following months from September to December, with 84 days. Since some

[1] http://www.b3.com.br/pt_br/market-data-e-indices/servicos-de-dados/market-data/historico/mercado-a-vista/cotacoes-historicas/

series present a higher proportion of one class over another in the training set, an oversampling technique was adopted to account for the class with fewer samples. The difference between the number of samples in both classes was calculated and from this difference repeated samples were randomly chosen to add the class with the lowest proportion. So for each different execution of the machine learning algorithm, a slightly different training set is obtained with the same proportion for each class, avoiding any bias in the model.

The architecture of the LSTM network used consists of 4 layers. The first layer contains 8 LSTM units, the second one contains 4 LSTM units and the third layer contains 2 LSTM units. The output layer contains a single neuron with a sigmoid activation function. Therefore, the network output is a value in the range $[0, 1]$ used to assign to one of the binary classes. The network is trained using *Adam* optimizer and binary cross-entropy loss function. The training is executed for 750 epochs with batches of size 100. This configuration was chosen after many tests, in which that configuration was found providing good generalization for the model, while at the same time, avoiding the overfitting problem.

The network input consists of 3 features: the values at time lags 1, 5, and 20 of each series. These features were chosen based on the variance ratio test performed in the previous step. So, for each day t of the training set, the neural network receives the 3 past values r_{t-19}, r_{t-4}, r_t, and the target r_{t+1} that will be a binary value of 0 (if r_{t+1} is negative) or 1 (if r_{t+1} is higher or equal to 0). Each row of the training set corresponds to these pairs of input and target data for each time t. It is expected that those series that showed significant correlations for these lags will provide better results in the classification model

For each time series, 50 executions of the neural network were performed and the arithmetic mean of the results was calculated. The neural network was implemented using the Python programming language as well as Keras neural network library.

E. Operation Strategy

Results of the prediction model provide a classification for each future day of the test set. Since it is used values related to the closing price, a strategy can be implemented that decides to buy or sell an stock from the future day's classification, minutes before the closing price of the current day. In this case, it is assumed that the value of the closing price of the current day will be roughly the same as at the time that the decision is made. Thus, the market operation strategy adopted was:

- High class: being an upward forecast, a purchase order is issued to the end of the current day t, and a sell order is placed at the end of the next day $t + 1$,
- Low Class: as a falling forecast, a sell order is issued at the end of the current day t, and a purchase order is placed at the end of the next day $t + 1$. Note that in this case the shorting operation is performed, since the investor did not previously own the stock.

It is noteworthy that by using the above strategy, there is a negotiation for each future day predicted by the learning algorithm in the previous step. Since the strategy adopted necessarily trades every day, even the days where there is no variation on the price should be included in one of the classes. So it was assumed to be included in the High Class (could also be included instead at the Low Class but will not make much difference in this context).

F. Analysis of Results

At this stage, the performance of the classification algorithm and the financial return at the end of market operations were analyzed. For the classification task, the results were compared with a random classifier which assigns to each future day one of the binary classes with equal probability, a Multilayer Perceptron (MLP) neural network with the same architecture of the LSTM, and a Support Vector Machine (SVM) with radial basis function kernel. The parameter gamma of the SVM was equal to the ratio of 1 and the number of features multiplied by the variance of the training set. And the regularization parameter (C) for each execution was chosen by a grid search between (0.1, 1, 5, 10, 15, 25, 50, 100, 150). It was used the *sklearn svc* package for implementation. The machine learn baselines used the same input features.

The metric chosen to evaluate the classification of the algorithms was the Accuracy since it measures the performance of both classes simultaneously. Accuracy is defined as the number of correct ratings on the total number of predictions:

$$A = \frac{TP + TN}{TP + FP + TN + FN}, \tag{10}$$

wherein *TP* is the true positive, *TN* is the true negative, *FP* is the false positive and *FN* is the false negative.

Concerning the financial return, it was also tested the Buy and Hold trading strategy, which consists of buying the stock and waiting until the end date. It was calculated the financial acumulated return at the end of the 84 trading period and compared the results from the different baselines with the LSTM.

IV. Results

This section is responsible for presenting the results found in the paper. Initially, Section A analyzes the statistical properties of the series. Then Section B analyzes the performance of the neural network against the baselines. Section C analyzes the strategy of operation of the classifiers in the financial market. Finally, Section D analyzes the operating cost of the adopted strategy.

A. Analysis of statistical properties of series

Table I shows the mean values, standard deviation, skewness, kurtosis and the p-value for the Jarque-Bera test. It can be seen that all stocks have a positive average very close to zero and with a standard deviation ranging from 0.01 to 0.04. The distributions tend to be symmetrical and the Jaque-Bera test showed that the majority of distributions are not normal.

TABLE I
FOUR FIRST MOMENTS OF THE SERIES AND JARQUE-BERA TEST

Stock	Mean	SD	Kurtosis	Skewness	JB
ABEV3	5e-4	0.01	3.00	0.16	0.66
BBAS3	3e-3	0.04	9.96	-0.41	2e-16
BRKM5	8e-4	0.03	7.08	0.12	2e-16
CPFE3	2e-3	0.01	3.87	0.53	0.008
EGIE3	1e-3	0.01	3.18	-0.25	0.29
ELET3	7e-3	0.03	3.79	0.58	0.01
JBSS3	1e-3	0.03	7.71	0.86	2e-16
LREN3	2e-3	0.02	2.89	-0.35	0.17
MULT3	2e-3	0.01	3.06	0.33	0.15
VIVT4	2e-3	0.02	3.23	-0.19	0.48
WEGE3	6e-4	0.02	23.87	-2.81	2e-16

This corroborates the stylized facts known in the literature that the return series do not have normal distributions, presenting heavy tails with high values of kurtosis [20]. When checking the value of stocks kurtosis, it is noted that non-normal distributions have higher values reaching 23.87 for WEGE3, demonstrating the presence of extreme values. Some of the series presented normal distributions, but the sample size was not large (166 days) and this fact may influence the behavior of the distribution. It is very likely that if a larger sample were used, for example, 2000 days, the probability distribution could be different.

Tables II, III, and IV show the R1, R2, S1, M1, and M2 statistics values for $N = 2$ (daily period), $N = 5$ (weekly period) and $N = 20$ (monthly period). In Table II, 5 stocks stand out: ABEV3, CPFE3, EGIE3, JBSS3 and WEGE3. All of these stocks presented significant statistics against RWH. In Table III, again the stocks CPF3, EGIE3 and JBSS3 presented results rejecting RWH. And in Table IV the stocks ABEV3 and CPFE3 again showed significant results. It is then noted that of the 11 stocks analyzed, 5 demonstrated to have a certain temporal dependence demonstrated by the different variance ratio tests while the remaining 6 stocks, for the tests performed, showed no rejection with RWH. These results suggest that applying forecasting models in these 5 highlighted series may yield better results than in the remaining others.

Using the 166 days, the critical values at the 2.5 % level of R1 are: -2,161, -2,020 and -1,768 for $N = 2$, $N = 5$ and $N = 20$ respectively. For R2 statistics, the values are equal to -2.145, -1.985 and -1.727. Finally, the critical values of S1 are equal to -2.017, -1.898 and -1.674.

B. LSTM neural network performance analysis

Figures 3, 4, 5 and 6 illustrate the boxplot graphs of Accuracy for 50 runs of the random classifier, LSTM network, MLP network, and SVM respectively. The 5 stocks highlighted before are shown in blue for easy distinction. The figures demonstrate that the distributions of the random classifier are very similar having the majority of values around 0.5, as expected. However, this behavior is not observed for machine learning algorithms, where several stocks are observed with distributions concentrated above 0.5. The 5 stocks that rejected the RWH hypothesis stand out for all cases. Table V shows the values for the average and standard deviation of the 50 runs of each classifier, for each stock. It is noted that except for the case of the random classifier, all stocks that rejected the RWH hypothesis had the highest values for the mean.

To compare the classifiers with each other, Welch's t-test was used, which is an adaptation of the Student's t-test. This test is used to compare the mean of samples from two groups when the variances are different and have the null hypothesis that the means are equal. Table VI shows the p-values obtained when comparing: random with each of the machine learning classifiers, LSTM with MLP and SVM, and finally in the last column, MLP with SVM. It is interesting to observe the comparisons of the algorithms with the random

TABLE II
VARIANCE RATIO TEST FOR $N = 2$

	N = 2				
Stock	**R1**	**R2**	**S1**	**M1**	**M2**
ABEV3	-2.22*	-1.81	-2.24*	-1.70	-1.75
BBAS3	0.79	1.01	-0.54	1.41	1.16
BRKM5	-1.82	-1.72	-1.16	-1.21	-1.09
CPFE3	-3.30*	-2.84*	-3.17*	-2.28*	-1.97*
EGIE3	-3.47*	-3.83*	-1.47	-3.76*	-3.40*
ELET3	0.10	1.02	-0.23	1.35	1.21
JBSS3	-3.09*	-3.19*	-2.08*	-2.96*	-2.98*
LREN3	-0.67	-0.60	-0.07	-0.70	-0.69
MULT3	-0.94	-0.86	-0.85	-0.69	-0.71
VIVT4	-1.69	-1.43	-1.16	-1.30	-1.34
WEGE3	-2.31*	-2.18*	-1.31	-1.31	-1.84

TABLE III
VARIANCE RATIO TEST FOR $N = 5$

	N = 5				
Stock	**R1**	**R2**	**S1**	**M1**	**M2**
ABEV3	-1.22	-0.79	-1.41	-0.62	-0.60
BBAS3	1.01	1.21	-0.39	0.86	0.58
BRKM5	-0.62	-0.54	-0.50	-0.58	-0.57
CPFE3	-2.80*	-2.72*	-2.48*	-2.46*	-2.14*
EGIE3	-2.50*	-2.56*	-1.52	-2.37*	-2.21*
ELET3	0.56	1.20	1.35	1.60	1.50
JBSS3	-2.15*	-2.51*	-0.62	-1.99*	-1.52
LREN3	-0.36	-0.40	0.22	-0.56	-0.56
MULT3	-1.35	-1.61	-0.73	-1.46	-1.52
VIVT4	-1.44	-1.33	-0.84	-1.28	-1.28
WEGE3	-1.29	-1.46	-1.18	-1.27	-1.32

TABLE IV
VARIANCE RATIO TEST FOR $N = 20$

	N = 20				
Stock	**R1**	**R2**	**S1**	**M1**	**M2**
ABEV3	-2.15*	-2.12*	-1.77*	-2.05*	-1.98*
BBAS3	0.97	1.12	0.39	0.29	0.20
BRKM5	-1.30	-1.25	-0.61	-1.07	-1.06
CPFE3	-1.66	-1.74*	-1.71*	-1.58	-1.37
EGIE3	-1.05	-1.00	-0.61	-0.91	-0.92
ELET3	-0.59	-0.25	0.50	-0.10	-0.09
JBSS3	-0.74	-1.16	1.46	-1.40	-1.23
LREN3	-0.46	-0.64	1.10	-0.67	-0.67
MULT3	-0.72	-1.24	0.32	-1.22	-1.26
VIVT4	-0.92	-0.94	-0.81	-0.98	-0.98
WEGE3	-0.72	-0.83	-0.85	-1.18	-1.25

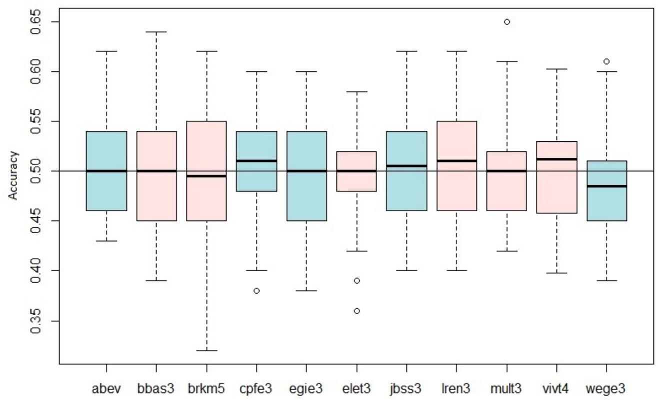

Fig. 3. Accuracy boxplot of the random classifier

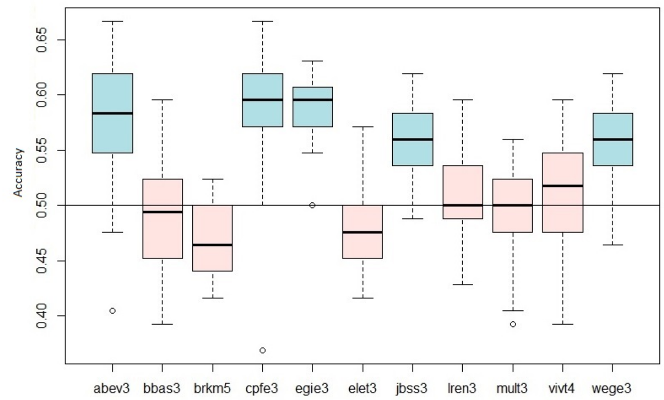

Fig. 5. Accuracy boxplot of the MLP classifier

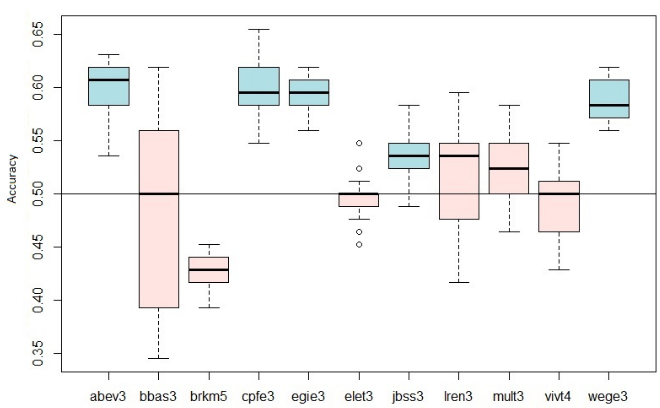

Fig. 4. Accuracy boxplot of the LSTM classifier

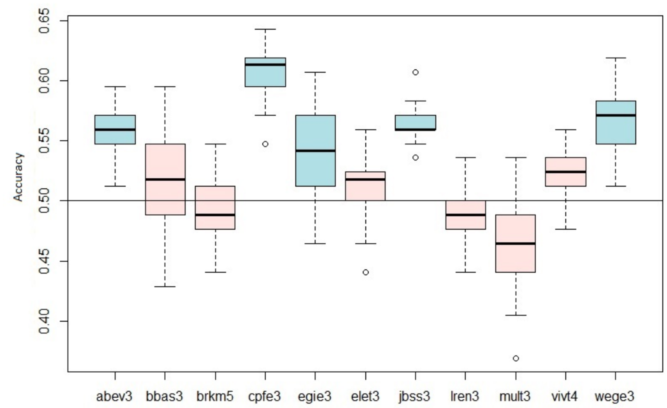

Fig. 6. Accuracy boxplot of the SVM classifier

since the results demonstrate that for the three algorithms, the 5 stocks that rejected the RWH hypothesis presented a statistically superior mean to the random, while the other 6 stocks presented a statistically inferior or equal mean to the random, except MULT3 in the case of LSTM and VIVT4 for SVM.

When comparing the three machine learning algorithms with each other, the LSTM stands out. LSTM presented an mean statistically higher than MLP for ABEV3 and WEGE3, an equivalent mean for CPFE3 and EGIE3 and a lower mean only for JBSS3. When compared to SVM, LSTM presented a higher mean for ABEV3, EGIE3 and WEGE3, an equivalent mean for CPFE3 and lower in the case of JBSS3. It is also interesting to observe the standard deviation, since it is ideal for a classifier to obtain a high mean concerning the Accuracy, and simultaneously a small value for the standard deviation. Figure 7 shows the values of the coefficient of variation (ratio of the standard deviation to the mean) obtained for each of the 5 stocks that stood out previously. The lower the value of the coefficient variation, the better the result. In the majority of cases, LSTM's superiority over baselines is again perceived.

C. Operation Strategy Analysis

Table VII shows the accumulated financial return values in percent at the end of the 84-day market operation for the period of test set. To compare the financial return achieved by the network, it also shows the results of the previously baselines using the same strategy and the Buy and Hold strategy for the same period. It is possible to see from the results of the Table the best performance of the LSTM network over the other baselines for 5 stocks that reject an RWH hypothesis in

TABLE V
VALUES OF THE AVERAGE AND STANDARD DEVIATION OF ACCURACY FOR THE RANDOM CLASSIFIER, LSTM, MLP AND SVM

	Random		LSTM		MLP		SVM	
Stock	**avg**	**sd**	**avg**	**sd**	**avg**	**sd**	**avg**	**sd**
ABEV3*	0.50	0.05	0.60	0.02	0.58	0.06	0.56	0.02
BBAS3	0.50	0.06	0.48	0.08	0.49	0.06	0.51	0.04
BRKM5	0.50	0.06	0.43	0.01	0.47	0.03	0.49	0.03
CPFE3*	0.51	0.05	0.60	0.03	0.57	0.09	0.60	0.03
EGIE3*	0.49	0.06	0.59	0.02	0.58	0.04	0.54	0.04
ELET3	0.50	0.05	0.50	0.02	0.48	0.03	0.51	0.03
JBSS3*	0.50	0.05	0.53	0.02	0.55	0.03	0.56	0.01
LREN3	0.51	0.06	0.52	0.04	0.51	0.03	0.49	0.02
MULT3	0.50	0.05	0.52	0.03	0.49	0.04	0.46	0.03
VIVT4	0.50	0.05	0.49	0.03	0.51	0.05	0.52	0.02
WEGE3*	0.50	0.06	0.59	0.02	0.56	0.04	0.57	0.03

TABLE VI
P-VALUES FOR THE WELCH'S T-TEST FOR ACURACCY MEANS BETWEEN CLASSIFIERS

	Random			LSTM		MLP
Stock	**LSTM**	**MLP**	**SVM**	**MLP**	**SVM**	**SVM**
ABEV3*	2.2e-16	8.1e-12	8.2e-13	1.3e-2	6.5e-15	5.2e-2
BBAS3	0.45	0.73	9.2e-2	0.62	4.1e-2	3.6e-2
BRKM5	5.7e-11	3.5e-3	0.55	1.6e-13	2.2e-16	1.2e-4
CPFE3*	2.2e-16	4.3e-6	2.2e-16	5.9e-2	0.31	2.0e-2
EGIE3*	2.2e-16	8.1e-15	1.9e-6	5.7e-2	8.5e-12	8.0e-7
ELET3	0.62	3.4e-2	6.1e-2	1.0e-2	2.2e-4	2.2e-6
JBSS3*	4.3e-4	1.3e-7	1.3e-10	9.7e-5	2.2e-16	5.1e-2
LREN3	0.11	0.85	2.0e-2	6.8e-2	2.5e-6	9.9e-5
MULT3	0.04	0.20	1.3e-5	4.2e-5	6.3e-15	1.7e-4
VIVT4	0.42	0.31	9.1e-3	0.03	2.8e-7	0.12
WEGE3*	6.4e-14	8.6e-8	9.5e-10	3.8e-5	2.6e-5	0.34

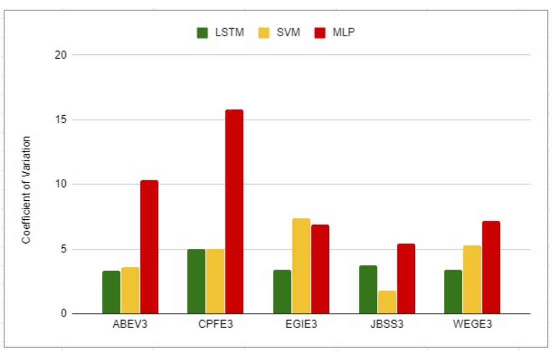

Fig. 7. Comparison of the coefficients of variation for ABEV3, CPFE3, EGIE3, JBSS3 and WEGE3 using LSTM, MLP, and SVM

the majority of the cases. These results also corroborate the hypothesis raised in the paper.

Figure 8 illustrates the comparison of LSTM's strategy performance for the 84 trading days between the three best and worst stocks, for one single simulation. It can be seen in the Figure the accumulated financial return over each day and the discrepancy in the final result of the accumulated financial return of the stocks that rejected RWH and those that did not.

An interesting fact that is observed is concerning the Accuracy and the financial return for different classifiers. In some cases, two algorithms showed very close values for the average Accuracy but with a very different financial return

TABLE VII
ACCUMULATED FINANCIAL RETURN (%) AT THE END OF 84 DAYS OF OPERATION

	Random	LSTM	MLP	SVM	BH
ABEV3*	0.48	13.72	18.44	11.18	-10.72
BBAS3	-5.27	9.26	-4.90	0.25	38.05
BRKM5	1.39	-8.63	6.43	29.00	29.68
CPFE3*	-0.87	5.54	4.93	5.28	6.01
EGIE3*	-1.91	17.71	15.86	7.49	-6.25
ELET3	-1.47	-2.98	13.86	35.23	9.50
JBSS3*	-5.15	10.16	1.76	-9.81	3.39
LREN3	0.73	7.32	16.00	-3.40	-2.43
MULT3	0.28	13.70	0.52	-17.90	5.11
VIVT4	0.09	-16.65	0.66	9.84	40.15
WEGE3*	-2.76	39.64	31.86	38.67	-9.57

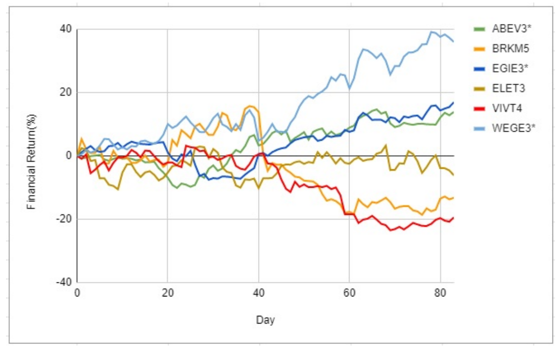

Fig. 8. Comparison of one simulation of the cumulative financial return over the 84 trading days. The chart shows the three best and worst stocks using LSTM strategy.

(for example LSTM and SVM for ELET3). Modeling the problem in discrete values of two classes (High and Low) ends up losing information in the magnitude of the original values. All values of the same class are considered equal and this can cause some problems, since it is known that these financial series do not have constant variance, [20]. Extreme values classified incorrectly can overcome several minor values classified correctly and in the end, a negative or unsatisfactory financial return can be obtained, even with a higher Accuracy.

D. Operating Cost Analysis

Even obtaining a positive financial return, it is necessary to evaluate the operational cost of each negotiation, since the strategy adopted uses daily operations and the total costs may exceed the profits depending on the size of the traded lot. For each transaction, 4 fees are charged: Income Tax withheld by B3(Brazilian Stock Exchange) of 0.5%, Brokerage Fee on transactions (R $ 2.50), Services Tax on Brokerage Fee (12 %) and Fees on the amount invested (0.025 %). If the investor has a profit on trading, the Income Tax paid by the investor on the profit (20 %) is also charged.

Table VIII presents the values relative to the average operating costs for ABEV3, CPFE3, EGIE3, JBSS3, and WEGE3, varying the negotiated lot size, using the predictions found by the LSTM neural network. The initial price for the first operating day was 19.22, 23.96, 38.53, 12.09 and 17.56 Reais

TABLE VIII
AVERAGE OPERATING COST VALUES, IN REAIS (R$), FOR DIFFERENT LOT SIZES L

	Operational Cost			
Stock	**L = 1**	**L = 100**	**L = 1000**	**L = 10000**
ABEV3	472.66	694.49	2730.56	23209.69
CPFE3	471.53	583.33	1600.47	11754.39
EGIE3	474.72	903.93	4802.51	43728.75
JBSS3	472.44	675.95	2508.63	20950.72
WEGE3	473.09	738.02	3153.16	27379.60

(R$) for the same stocks. Comparing the operational costs by lot with the initial price and accumulated financial return (Table VII) it is possible to see some interesting results.

For lots of up to 100 stocks, the trading strategy adopted is not interesting since the gross profit values are lower than the operating cost. Since the strategy uses two operations per day, the fixed cost of these operations overlaps with the amount of profit made for small lots. The strategy begins to become viable with lots of 1000 stocks, as can be seen by checking the values of financial return and operating cost of stocks EGIE3, and WEGE3 for example. However, even using lots of 10,000 stocks and having a positive financial return, there are cases where operating cost overrides profit, for example for the JBSS3 stock. These results demonstrate the complexity to create efficient strategies: even obtaining positive financial returns it is necessary to use a large volume of stocks to cover the operational costs.

V. CONCLUSION

In this paper, we studied the statistical properties of 11 stocks of the Brazilian market, and the relationship of these properties with the predictability of the series using an LSTM neural network for classification. The results obtained by the experiments were promising, with good performance in classification and also financial return, compared to the baselines.

A methodology was developed based on the hypothesis that it is possible to choose some specific stocks to have good forecasting results by applying the variance ratio tests in the data training set. The paper showed evidence that there is possibly a relationship between the series that stand out in these statistical tests and the performance of the neural network. The 5 stocks that rejected the RWH in the tests obtained the best results in terms of classification among all others. Also, the other 6 stocks that did not reject the RWH showed poor results with in some cases the random classifier outperformed the machine learning algorithms. The same behavior was also observed in the results regarding the financial return.

As future works, we intend to use the same methodology in different stock series to confirm the possible relationship between the variance ratio tests and the predictability of the series when using machine learning algorithms. It is also ideal to test for different periods for more robust results. It is important to notice that in the literature there are several different tests of RWH that exploit different proprieties besides the variance of the series. These different tests could also be analyzed with combination with machine learning algorithms to test the predictability of the series.

Another aspect for future work is to model the classification problem in 3 classes: upward, downward and neutral trends. This modeling can increase the performance of trading strategies since in many moments financial series behave with low variation and it is ideal to avoid negotiation in these neutral moments.

REFERENCES

[1] L. Bachelier, ”Theory of speculation,” Reprinted in P. Cootner(ed.), 1964, The Character of Stock Market Prices, MIT Press, Cambridge, Massachusetts, 1900, pp. 17-78

[2] Lo, Andrew W and MacKinlay, A. Craig, “Stock Market Prices Do Not Follow Random Walks: Evidence From a Simple Specification Test,”National Bureau of Economic Research, 1988

[3] Chow, K. V., Denning, K. C.”A simple multiple variance ratio test,” Journal of Econometrics, pp. 385–401, 1993

[4] Wright, J. H.”Alternative variance-ratio tests using ranks and signs,” Journal of Business & Economic Statistics, pp. 1-9, 2000

[5] Whang, Y. J., Kim, J.”A multiple variance ratio test using subsampling,” Economics Letters, pp. 225–230, 2003

[6] Hafiz A.A.B. Hoque and Jae H. Kim and Chong Soo Pyun ”A comparison of variance ratio tests of random walk: A case of Asian emerging stock markets,” International Review of Economics & Finance, pp. 488-502, 2007

[7] Jae H. Kim and Abul Shamsuddin ”Are Asian stock markets efficient? Evidence from new multiple variance ratio tests,” Journal of Empirical Finance, pp. 518-532, 2008

[8] D. M. Q. Nelson and A. C. M. Pereira and R. A. de Oliveira “Stock market’s price movement prediction with LSTM neural networks,”Neural 2017 International Joint Conference on Neural Networks (IJCNN), pp. 1419-1426, 2017

[9] Huang, Biao and Ding, Qiao and Sun, Guozi and Li, Huakang“Stock Prediction Based on Bayesian-LSTM,”Proceedings of the 2018 10th International Conference on Machine Learning and Computing, pp. 128-133, 2018

[10] Lin, Moule and Chen, Changxi “Short-term Prediction of Stock Market Price Based on GA Optimization LSTM Neurons,”Proceedings of the 2018 2Nd International Conference on Deep Learning Technologies, pp. 66–70, 2018

[11] Pawar, Kriti“Stock Market Price Prediction Using LSTM RNN,”Emerging Trends in Expert Applications and Security, pp. 493-503, 2019

[12] A. J. P. Samarawickrama and T. G. I. Fernando“A recurrent neural network approach in predicting daily stock prices an application to the Sri Lankan stock market,”IEEE International Conference on Industrial and Information Systems (ICIIS), pp. 1-6, 2017

[13] K. Chen and Y. Zhou and F. Dai“A LSTM-based method for stock returns prediction: A case study of China stock market,”IEEE International Conference on Big Data (Big Data), pp. 2823-2824, 2015

[14] Z. Zhao and R. Rao and S. Tu and J. Shi“Time-Weighted LSTM Model with Redefined Labeling for Stock Trend Prediction,”IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI), pp. 1210-1217, 2017

[15] S. Yao and L. Luo and H. Peng“High-Frequency Stock Trend Forecast Using LSTM Model,”13th International Conference on Computer Science Education (ICCSE), pp. 1-4, 2018

[16] X. Shao and D. Ma and Y. Liu and Q. Yin“Short-term forecast of stock price of multi-branch LSTM based on K-means,”4th International Conference on Systems and Informatics (ICSAI), pp. 1546-1551, 2017

[17] Stephen J Taylor, “Asset Price Dynamics, Volatility and Prediction,”Princeton University Press, pp. 102-103, 2005

[18] Hochreiter, Sepp and Schmidhuber, Jürgen, “Long short-term memory,”Neural computation, pp. 1735-1780, 1997

[19] Olah C. (2019) ”Understanding LSTM Networks” https://colah.github.io/posts/2015-08-Understanding-LSTMs/, Accessed 15 December 2019

[20] Cont, R.”Empirical properties of asset returns: stylized facts and statistical issues,” Quantitative Finance, pp. 223-236, 2001